NONPARAMETRIC PRIORS ON COMPLETE SEPARABLE METRIC SPACES

By Peter Orbanz

University of Cambridge

A Bayesian model is nonparametric if its parameter space has infinite dimension; typical choices are spaces of discrete measures and Hilbert spaces. We consider the construction of nonparametric priors when the parameter takes values in a more general functional space. We (i) give a Prokhorov-type representation result for nonparametric Bayesian models; (ii) show how certain tractability properties of the nonparametric posterior can be guaranteed by construction; and (iii) provide an ergodic decomposition theorem to characterize conditional independence when de Finetti's theorem is not applicable. Our work is motivated primarily by statistical problems where observations do not form exchangeable sequences, but rather aggregate into some other type of random discrete structure. We demonstrate applications to two such problems, permutation-valued and graph-valued observations, and relate our results to recent work in discrete analysis and ergodic theory.

1. Introduction. Nonparametric priors can be classified according to the type of random functional which serves as a model parameter: Examples include discrete measures [18, 19, 23, 27], continuous functions [50], mixtures and other density functions [25, Chapter 5], and monotone functions [26, Chapter 3]. Bayesian nonparametric statistics predominantly revolves around various mathematical modeling primitives available on each of these domains, such as the Dirichlet process (DP) [19, 23] and the Gaussian process (GP) [1, 10, 47]. Adaptations of these modeling primitives to the needs of specific applications, and the modular way in which they can be combined to express hierarchical structures, account for the vast majority of the large and growing literature on nonparametric priors [26, 49]. The rather limited attention which certain important types of data—e.g. networks, relational data, or ranked lists—have to date received in the field is arguably due to a lack of readily available primitives. The design of nonparametric priors for such problems raises a range of mathematical questions. A small subset of these are addressed in this paper.

AMS 2000 subject classifications: Primary 62C10, 62G05

Keywords and phrases: Bayesian nonparametrics, random measures, ergodic decomposition, Dirichlet process, Gaussian process, exchangeability, projective limits

Nonparametric priors tend to be technically more challenging than their parametric counterparts since the underlying parameter space is not locally compact; prior and posterior hence have no density representation with respect to a translation-invariant measure. Additionally, Bayes' theorem is often not applicable, since a nonparametric posterior is not generally dominated [45]. A cornerstone of nonparametric constructions are therefore representations which substitute for densities and permit the computation of posteriors. Most representations used in the literature—stick-breaking, Lévy processes, etc—are too model-specific for the purposes of generalization.

We build on work of Lauritzen [36] who, in a different context, introduced projective limits of regular conditional probabilities which he referred to as *projective statistical fields*. Much like the defining components of a nonparametric Bayesian model, his limit objects are regular conditional probabilities on a space of infinite dimension. Projective statistical fields thus almost provide a generic representation of nonparametric priors and posteriors. Not quite, since random quantities arising as parameters in Bayesian nonparametrics exhibit almost sure regularity properties—e.g. continuity or σ-additivity—which projective limits do not express. The missing link between projective limit representations and regularity are "Prokhorov conditions" for stochastic processes, i.e. tightness conditions on compact sets [e.g. 12, Chapter III]. Using such regularity conditions, we obtain a construction which, roughly speaking, represents a nonparametric Bayesian model by a projective system of parametric Bayesian models (Theorems 2.3 and 2.5).

With this representation in place, we derive conditions under which the existence of analytic update formulae for posterior distributions is guaranteed (Theorem 2.7). In the specific case of nonparametric models obtained as limits of exponential families, we obtain an explicit formula (Corollary 2.8). For types of data not representable as exchangeable sequences, the requisite conditional independence properties of a Bayesian model cannot be identified, as they usually are, by de Finetti's theorem. For such problems, de Finetti's result can be substituted by more general ergodic decomposition theorems [15, 29, 36]. We give a version of such a theorem tailored to Bayesian nonparametrics (Theorem 2.9), which also establishes a close connection between popular nonparametric priors and Lauritzen's work on extremal families [36].

We consider applications to two types of data; observations which aggregate into dense graphs or into ranked lists (permutations). Analytic properties of the sample spaces are given in recent work from Lovász and Szegedy [38, 39] for dense graphs, and from Kerov, Olshanski, and Vershik [33] for permutations. The derivation of Bayesian statistical models using our results is demonstrated for both problems in some detail. To clarify the connection to existing methods in the Bayesian nonparametric literature, we also show how the familiar Gaussian process and Dirichlet process priors arise as instances of our approach.

Article Structure. Results are stated in Sec. 2 and illustrated by applications in Sec. 3. Related work is discussed in Sec. 4. All proofs are collected in App. A, and App. B summarizes relevant facts on projective limits.

Notation. We assume throughout that the underlying model of randomness is an abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If $\mathbf{X}$ is a topological space, $\mathbf{M}(\mathbf{X})$ denotes the set of probability measures on the Borel sets $\mathcal{B}(\mathbf{X})$. For a mapping $\phi : \mathbf{X} \to \mathbf{X}'$, we denote the push-forward (the image measure) of a measure $\mu \in \mathbf{M}(\mathbf{X})$ as $\phi_{\#}\mu \in \mathbf{M}(\mathbf{X}')$. Similarly, $\phi^{\#}\nu \in \mathbf{M}(\mathbf{X})$ is the pull-back of $\nu \in \mathbf{M}(\mathbf{X}')$, i.e. the measure defined implicitly by $\phi_{\#}(\phi^{\#}\nu) = \nu$. We write ν^* for the outer measure of ν. The pull-back exists if $\nu^*(\phi(\mathbf{X})) = \nu(\mathbf{X}')$ [22, 234F]. $\mathrm{Law}(X) := X_{\#}\mathbb{P}$ denotes the law of a random variable $X : \Omega \to \mathbf{X}$. We use the term *probability kernel* for measurable mappings of the form $\mathbf{p} : \mathbf{X}'' \to \mathbf{M}(\mathbf{X})$. A kernel $\mathbf{p}$ is also denoted as a function $\mathbf{p}(A, x'')$ on $\mathcal{B}(\mathbf{X}) \times \mathbf{X}''$, and its push-forward is $(\phi_{\#}\mathbf{p})(\,.\,, x'') := \phi_{\#}\mathbf{p}(\,.\,, x'')$.

2. Results.

We define a *Bayesian model* as a random probability measure Π on a sample space $\mathbf{X}$, i.e. a probability kernel $\Pi : \Omega \to \mathbf{M}(\mathbf{X})$. We generally decompose Π as follows: Let $\mathbf{T}$ be another space, called the *parameter space* of the model. We assume that there is a probability kernel $\mathbf{p} : \mathbf{T} \to \mathbf{M}(\mathbf{X})$, called the *likelihood*, and a random variable $\Theta : \Omega \to \mathbf{T}$, called the *parameter*, such that $\Pi = \mathbf{p} \circ \Theta$. Observations are described by a random variable $X : \Omega \to \mathbf{X}$ with conditional probability $\mathbf{p}$ given Θ. The law $Q := \Theta_{\#}\mathbb{P}$ of the parameter is the *prior distribution* of the model. A probability kernel $\widehat{\mathbf{q}} : \mathbf{X} \to \mathbf{M}(\mathbf{T})$ given by $\mathbb{P}[\Theta \in \,.\,|X = x] =_{\text{a.s.}} \widehat{\mathbf{q}}(\,.\,, x)$ is the *posterior* under observation x [21, 45].

In Bayesian statistics, the prior distribution Q is usually itself parameterized by a *hyperparameter* variable Y with values $y \in \mathbf{Y}$. The resulting family of priors is then a probability kernel $\mathbf{q}(\,.\,, y) = \mathrm{Law}(\Theta^y) = \mathbb{P}[\Theta \in \,.\,|Y = y]$, and the posterior takes the form $\widehat{\mathbf{q}} : \mathbf{X} \times \mathbf{Y} \to \mathbf{M}(\mathbf{T})$. We refer to a Bayesian model as *nonparametric* if the parameter space $\mathbf{T}$ has infinite dimension, and as *parametric* if the dimension is finite. The main analytical implication is that $\mathbf{T}$ is not locally compact if the model is nonparametric.

2.1. *Representations.*

The presentation in this section assumes basic familiarity with projective limits of measures [8, 9]. A summary of relevant definitions is given in App. B. To define statistical models on infinite-dimensional spaces $\mathbf{T}$ and $\mathbf{X}$, the spaces are represented by means of projective limits. More precisely, to represent a space $\mathbf{X}$, a family $\langle \mathbf{X}_{\mathrm{I}} \rangle_{\Gamma}$ of finite-dimensional spaces $\mathbf{X}_{\mathrm{I}}$ is assembled into a projective limit space $\mathbf{X}_{\Gamma} := \varprojlim \langle \mathbf{X}_{\mathrm{I}} \rangle_{\Gamma}$. Since the spaces of "regular" mappings of interest in Bayesian nonparametrics—for instance, the sets $\mathbf{X} = C(\mathbb{R}_+, \mathbb{R})$ of continuous functions, $\mathbf{X} = \mathbf{M}(\mathbb{R})$ of probability measures or $\mathbf{X} = D(\mathbb{R}_+, \mathbb{R})$ of cádlág functions—are not directly representable as a projective limit $\mathbf{X}_{\Gamma}$, we represent $\mathbf{X}$ indirectly by embedding it into $\mathbf{X}_{\Gamma}$. Conditional probabilities $\mathbf{p}$ on $\mathbf{X}$ are then represented by families of conditional probabilities $\mathbf{p}_{\mathrm{I}}$ on $\mathbf{X}_{\mathrm{I}}$.

Some of the complete separable metric spaces of interest in the following admit more than one relevant compatible metric, and we hence state results in terms of spaces which are *Polish*, i.e. complete, separable and metrizable. Let $\langle \mathbf{X}_{\mathrm{I}} \rangle_{\Gamma}$ be a family of Polish spaces indexed by a partially ordered, directed index set $(\Gamma, \preceq)$. Throughout, Γ is countable. We assume that the spaces $\mathbf{X}_{\mathrm{I}}$ are linked by generalized projections $f_{\mathrm{JI}} \colon \mathbf{X}_{\mathrm{J}} \to \mathbf{X}_{\mathrm{I}}$ whenever $I \preceq J$, and that the mappings f_{JI} are continuous and surjective. The projective limit space is denoted $\mathbf{X}_{\Gamma} := \varprojlim \langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$, with canonical mappings $f_{\mathrm{I}} : \mathbf{X}_{\Gamma} \to \mathbf{X}_{\mathrm{I}}$ (cf. App. B). Since Γ is countable and the f_{JI} are surjective, $\mathbf{X}_{\Gamma}$ is again Polish [30, §17D] and non-empty [8, III.7.4, Proposition 5].

DEFINITION 2.1. A topological space $\mathbf{X}$ is *embedded into a projective system* $\langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$ of Polish spaces if there are continuous, surjective mappings $\mathrm{F}_{\mathrm{I}} : \mathbf{X} \to \mathbf{X}_{\mathrm{I}}$ for all $I \in \Gamma$ satisfying the conditions

$$f_{\mathrm{JI}} \circ \mathrm{F}_{\mathrm{J}} = \mathrm{F}_{\mathrm{I}} \quad \text{if } I \preceq J \text{ and } \quad \forall x \neq x' \; \exists I \in \Gamma : \quad \mathrm{F}_{\mathrm{I}}(x) \neq \mathrm{F}_{\mathrm{I}}(x') \; . \tag{2.1}$$

The projective limit of the mappings $\langle \mathrm{F}_{\mathrm{I}} \rangle_{\Gamma}$ is denoted $\mathrm{F} := \varprojlim \langle \mathrm{F}_{\mathrm{I}} \rangle_{\Gamma}$. As a projective limit of continuous mappings, $\mathrm{F} : \mathbf{X} \to \mathbf{X}_{\Gamma}$ is continuous and satisfies $\mathrm{F}_{\mathrm{I}} = f_{\mathrm{I}} \circ \mathrm{F}$, but it is *not* generally surjective. By condition (2.1), F is injective, and hence embeds $\mathbf{X}$ into $\mathbf{X}_{\Gamma}$. The individual components form a commutative diagram:

$$\begin{tikzcd}
& \mathbf{X} \arrow[d, "\mathrm{F}"] \arrow[ddl, bend right, "\mathrm{F}_{\mathrm{J}}"'] \arrow[ddr, bend left, "\mathrm{F}_{\mathrm{I}}"] & \\
& \mathbf{X}_{\Gamma} \arrow[dl, "f_{\mathrm{J}}"'] \arrow[dr, "f_{\mathrm{I}}"] & \\
\mathbf{X}_{\mathrm{J}} \arrow[rr, "f_{\mathrm{JI}}"] & & \mathbf{X}_{\mathrm{I}}
\end{tikzcd} \tag{2.2}$$

In the statistical models considered further on, $I \in \Gamma$ describes the dimension of parameter space. As the number of observations grows, a larger number of parameter dimensions is required to explain the sample, and I increases in terms of $\preceq$.

A family $\langle P_{\mathrm{I}} \rangle_{\Gamma}$ of probability measures P_{I} on the spaces $\mathbf{X}_{\mathrm{I}}$ is called a *promeasure* (or *projective family*) if $f_{\mathrm{JI}\#}P_{\mathrm{J}} = P_{\mathrm{I}}$ whenever $I \preceq J$, i.e. if each measure P_{I} is the marginal of all higher-dimensional measures P_{J} in the family [9]. We define an analogous concept for families of regular conditional probabilities $\mathbf{p}_{\mathrm{I}}(\,.\,,\omega) = \mathbb{P}[X_{\mathrm{I}} \in \,.\,|\mathcal{C}_{\mathrm{I}}](\omega)$.

Definition 2.2. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A family $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ of probability kernels $\mathbf{p}_{\mathrm{I}} : \Omega \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ is called a *conditional promeasure* if

$$f_{\mathrm{JI}\#}\mathbf{p}_{\mathrm{J}} =_{\mathrm{a.s.}} \mathbf{p}_{\mathrm{I}} \quad \text{whenever } I \preceq J\ . \tag{2.3}$$

It is *tight on* $\mathbf{X}$ if, for each $\varepsilon > 0$, there exist compact sets $K^{\omega} \subset \mathbf{X}$ with

$$\mathbf{p}_{\mathrm{I}}(\mathrm{F}_{\mathrm{I}}K^{\omega}, \omega) > 1 - \varepsilon \qquad \mathbb{P}\text{-a.s. for all } I \in \Gamma\ . \tag{2.4}$$

Analogously, a promeasure $\langle P_{\mathrm{I}} \rangle_{\Gamma}$ is tight on $\mathbf{X}$ if for each $\varepsilon > 0$ there is a compact $K \subset \mathbf{X}$ such that $P_{\mathrm{I}}(\mathrm{F}_{\mathrm{I}}K) > 1 - \varepsilon$ for all I.

A tight promeasure $\langle P_{\mathrm{I}} \rangle_{\Gamma}$ uniquely defines a probability measure P on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$, by Prokhorov's extension theorem [9, IX.4.2, Theorem I]. Conversely, any measure P on $\mathbf{X}$ defines a tight promeasure $\langle \mathrm{F}_{\mathrm{I}\#}P \rangle_{\Gamma}$. This bijective correspondence makes families of finite-dimensional measures useful proxies for stochastic processes. Our first result generalizes the correspondence from measures to the conditional probabilities relevant to Bayesian statistics. For a family of σ-algebras $\mathcal{C}_{\mathrm{I}} \subset \mathcal{A}$, we denote the limit σ-algebra by $\mathcal{C}_{\Gamma} := \sigma(\mathcal{C}_{\mathrm{I}}; I \in \Gamma)$ and the tail σ-algebra by $\mathcal{C}_{\mathrm{T}} := \limsup_{\mathrm{I}\in\Gamma} \mathcal{C}_{\mathrm{I}}$.

Theorem 2.3. *Let* $\mathbf{X}$ *be a Hausdorff space embedded into* $\langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$ *as in* (2.2)*, and let* $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ *be a conditional promeasure of kernels* $\mathbf{p}_{\mathrm{I}} : \Omega \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$*. There is a probability kernel* $\mathbf{p} \colon \Omega \to \mathbf{M}(\mathbf{X})$ *satisfying*

$$\mathrm{F}_{\mathrm{I}\#}\mathbf{p} =_{a.s.} \mathbf{p}_{\mathrm{I}} \qquad \textit{for all } I \in \Gamma \tag{2.5}$$

if and only if $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ *is tight on* $\mathbf{X}$*. The conditional promeasure uniquely determines* $\mathbf{p}$ *outside a* $\mathbb{P}$*-null set, and* $\mathbf{p}$ *almost surely takes values in the set* $\mathbf{RM}(\mathbf{X}) \subset \mathbf{M}(\mathbf{X})$ *of Radon probability measures. If* $\mathbf{p}_{\mathrm{I}}$ *is* $\mathcal{C}_{\mathrm{I}}$*-measurable for each* I*, then* $\mathbf{p}$ *has both a* $\mathcal{C}_{\Gamma}$*-measurable and a* $\mathcal{C}_{\mathrm{T}}$*-measurable version.*

The limit object is denoted $\mathrm{F}^{\#} \varprojlim \langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma} := \mathbf{p}$. The notation is motivated by the fact that $\mathbf{p}$ can be regarded as a pull-back: A kernel $\mathbf{p}_{\Gamma} : \Omega \to \mathbf{M}(\mathbf{X}_{\Gamma})$

can be constructed by setting $\mathbf{X} := \mathbf{X}_\Gamma$ in the theorem. In this case, tightness is trivial: Any conditional promeasure is tight on $\mathbf{X}_\Gamma$. If $\langle \mathbf{p}_{\mathrm{I}} \rangle_\Gamma$ is additionally tight on some other, embedded space $\mathbf{X}$, Lemma A.2 shows that the limit $\mathbf{p}$ on this space is a version of the pull-back $\mathrm{F}^{\#}\mathbf{p}_\Gamma$. In the parlance of stochastic process theory, $\mathbf{p}(\,.\,,\omega)$ is a modification of $\mathbf{p}_\Gamma(\,.\,,\omega)$ with paths in $\mathbf{X}$.

For our purposes, the probability kernels $\mathbf{p}_{\mathrm{I}}$ are typically given by parametric models, in which case each $\mathbf{p}_{\mathrm{I}}$ is defined on a separate parameter space $\mathbf{T}_{\mathrm{I}}$. In this case, Theorem 2.3 immediately translates into:

Corollary 2.4. *Let $\mathbf{T}$ be a Polish space, embedded into $\langle \mathbf{T}_{\mathrm{I}}, g_{\mathrm{JI}} \rangle_\Gamma$ by maps $\langle \mathrm{G}_{\mathrm{I}} \rangle_\Gamma$, and Q a probability measure on $\mathcal{B}(\mathbf{T})$. For probability kernels $\mathbf{p}_{\mathrm{I}} : \mathbf{T}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$, the family $\langle \mathbf{p}_{\mathrm{I}} \circ \mathrm{G}_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure defined on $(\mathbf{T}, \mathcal{B}(\mathbf{T}), Q)$ iff $f_{\mathrm{JI}\#}\mathbf{p}_{\mathrm{J}} =_{a.s.} \mathbf{p}_{\mathrm{I}} \circ g_{\mathrm{JI}}$ for $I \preceq J$. If it satisfies (2.4), there is an a.s.-unique kernel $\mathbf{p} : \mathbf{T} \to \mathbf{M}(\mathbf{X})$ satisfying $\mathrm{F}_{\mathrm{I}\#}\mathbf{p} =_{a.s.} \mathbf{p}_{\mathrm{I}} \circ \mathrm{G}_{\mathrm{I}}$ for $I \in \Gamma$.*

The Hausdorff space formulation in Theorem 2.3 is more general than typically required for Bayesian models, whose natural habitat are Polish spaces. If $\mathbf{X}$ is Polish, all conditional probabilities have a regular version and $\mathbf{RM}(\mathbf{X}) = \mathbf{M}(\mathbf{X})$. It is worth noting, though, that regularity of the conditional probabilities $\mathbf{p}_{\mathrm{I}}$ and the Polish topologies on the spaces $\mathbf{X}_{\mathrm{I}}$ ensure $\mathbf{p}$ to be a regular conditional probability and almost surely Radon, even if $\mathbf{X}$ itself is not Polish. We will from here on restrict our attention to the Polish case, but we note that there are examples in Bayesian nonparametrics which, at least in principle, require less topological structure. Sethuraman's representation of the Dirichlet process—for which the domain V of the Dirichlet random measure, and hence the parameter space $\mathbf{T} = \mathbf{M}(V)$, are not required to be Polish—is a case in point [23].

2.2. *Construction of nonparametric Bayesian models.* Let $\mathbf{X}$ be embedded as above, and let $\mathbf{T}$ be a parameter space embedded into $\langle \mathbf{T}_{\mathrm{I}}, g_{\mathrm{JI}} \rangle_\Gamma$ by $\langle \mathrm{G}_{\mathrm{I}} \rangle_\Gamma$, where we generally assume $\mathbf{T}_{\mathrm{I}}$ to be of finite dimension. Let $\mathbf{p}_{\mathrm{I}} \colon \mathbf{T}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ and $\mathbf{q}_{\mathrm{I}} \colon \mathbf{Y} \to \mathbf{M}(\mathbf{T}_{\mathrm{I}})$ be probability kernels and $\Theta^y_{\mathrm{I}} \colon \Omega \to \mathbf{T}_{\mathrm{I}}$ a random variable with distribution $\mathbf{q}_{\mathrm{I}}(\,.\,,y)$. Then $\Pi^y_{\mathrm{I}} = \mathbf{p}_{\mathrm{I}} \circ \Theta^y_{\mathrm{I}}$ is a parametric Bayesian model on $\mathbf{X}_{\mathrm{I}}$, with a posterior $\widehat{\mathbf{q}}_{\mathrm{I}}(\,.\,,x,y) = \widehat{\mathbf{q}}^y_{\mathrm{I}}(\,.\,,x)$. If the families $\langle \mathbf{p}_{\mathrm{I}} \rangle_\Gamma$ and $\langle \mathbf{q}_{\mathrm{I}} \rangle_\Gamma$ have well-defined limits on $\mathbf{X}$ and $\mathbf{T}$, they define a nonparametric Bayesian model with posterior $\widehat{\mathbf{q}}$ and parameter space $\mathbf{T}$, but in a non-constructive sense, since the posterior is a purely abstract quantity.

However, parametric models are typically well-parameterized in the sense that the observation variable X_{J} on each space $\mathbf{X}_{\mathrm{J}}$ depends on Θ_{I} only through X_{I} if $I \preceq J$. In this case, the posteriors form a conditional promeasure (Lemma A.1), and the model becomes much more tractable.

THEOREM 2.5. *Let $\langle \mathbf{p}_{\mathrm{I}} \circ \mathrm{G}_{\mathrm{I}} \rangle_{\Gamma}$ and $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ be tight conditional promeasures on Polish spaces $\mathbf{X}$ and $\mathbf{T}$, respectively, with limits $\mathbf{p}$ and $\mathbf{q}$. If the family $\langle \widehat{\mathbf{q}}_{\mathrm{I}} \rangle_{\Gamma}$ of posteriors also forms a conditional promeasure, the following holds:*
(i) *The family $\langle \widehat{\mathbf{q}}_{\mathrm{I}} \rangle_{\Gamma}$ is tight on $\mathbf{T}$ and $\mathrm{G}^{\#} \varprojlim \langle \widehat{\mathbf{q}}_{\mathrm{I}} \circ \mathrm{F}_{\mathrm{I}} \rangle_{\Gamma} =_{a.s.} \widehat{\mathbf{q}}$.*
(ii) *Let $(I_n)_n$ be a cofinal sequence in Γ. Fix $y \in \mathbf{Y}$ and a sequence $(x_{\mathrm{I}_n})_n$ of points $x_{\mathrm{I}_n} \in \mathbf{X}_{\mathrm{I}_n}$ satisfying $f_{\mathrm{I}_{n+1}\mathrm{I}_n} x_{\mathrm{I}_{n+1}} = x_{\mathrm{I}_n}$ for all n. Then for any sequence $(z^{(\mathrm{n})})_n$ in $\mathbf{X}$ with $\mathrm{F}_{\mathrm{I}_n} z^{(\mathrm{n})} = x_{\mathrm{I}_n}$, the posterior on $\mathbf{T}$ converges weakly,*

$$\widehat{\mathbf{q}}^{y}(\,.\,, z^{(\mathrm{n})}) \quad \xrightarrow{\;w\;} \quad \widehat{\mathbf{q}}^{y}(\,.\,, \mathrm{F}^{-1} \varprojlim \langle x_{\mathrm{I}_n} \rangle_{\mathbb{N}}) \qquad X_{\#}\mathbb{P}\text{-}a.s. \text{ on } \mathbf{X}\,. \tag{2.6}$$

In other words, there is a random element X in $\mathbf{X}$ such that the conditional probability of $\mathrm{F}_{\mathrm{I}} X$ given Θ is precisely $\mathbf{p}_{\mathrm{I}}$. This element is of infinite dimension and never fully observed; rather, n observation steps yield "censored" information $X_{\mathrm{I}_n} = \mathrm{F}_{\mathrm{I}_n} X$. In the simplest case, X_{I_n} is a sequence of length n. More generally, it is a function of the first n recorded observations, for instance, each observation may be a vertex in a graph and its associated edges, and X_{I_n} the aggregate subgraph obtained after n steps. Theorem 2.5(i) shows that the diagram

$$\begin{tikzcd}[column sep=huge] \mathbf{q} \arrow[r, "\{X = x\}"] \arrow[d, shift right, "\mathrm{G}_{\mathrm{I}}"'] & \widehat{\mathbf{q}} \arrow[d, shift right, "\mathrm{G}_{\mathrm{I}}"'] \\ \mathbf{q}_{\mathrm{I}} \arrow[r, "\{X_{\mathrm{I}} = \mathrm{F}_{\mathrm{I}} x\}"] \arrow[u, shift right, "\mathrm{G}^{\#}\varprojlim"'] & \widehat{\mathbf{q}}_{\mathrm{I}} \arrow[u, shift right, "\mathrm{G}^{\#}\varprojlim"'] \end{tikzcd} \tag{2.7}$$

commutes. If each censored observation is augmented arbitrarily to form a full element $z^{(\mathrm{n})}$ of $\mathbf{X}$, part (ii) shows that posteriors computed from $z^{(\mathrm{n})}$ converge weakly to the actual posterior. Convergence holds almost surely under the prior $\mathbf{q}(\,.\,, y)$. All caveats pertaining results which neglect null sets under the prior apply [e.g. 26, Chapter 2].

The models $(\mathbf{p}_{\mathrm{I}}, \mathbf{q}_{\mathrm{I}})$ used in the construction will in most cases have separate, finite-dimensional hyperparameter spaces $\mathbf{Y}_{\mathrm{I}}$. It is then useful to collect these spaces in a projective system $\langle \mathbf{Y}_{\mathrm{I}}, h_{\mathrm{JI}} \rangle_{\Gamma}$, into which $\mathbf{Y}$ is embedded by mappings $\langle \mathrm{H}_{\mathrm{I}} \rangle_{\Gamma}$.

EXAMPLE 2.6. We illustrate the construction by a familiar model: Choose $\mathbf{T}$ as the Hilbert space $\mathbf{T} = L_2[0,1]$. Let Γ consist of all finite subsets of $\mathbb{N}$, ordered by inclusion, and define $\mathbf{T}_{\mathrm{I}} = \mathbf{Y}_{\mathrm{I}} = \mathbb{R}^{\mathrm{I}}$. The projective limits space $\mathbf{T}_{\Gamma} = \mathbb{R}^{\mathbb{N}}$ contains ℓ_2 as a proper subset. Hence, the Hilbert space isomorphism $\mathrm{G} : L_2[0,1] \to \ell_2$ embeds $L_2[0,1]$ into $\mathbf{T}_{\Gamma}$. Choose $\mathbf{Y}$ identically as $\mathbf{Y} := \mathbf{T}$ with $\mathrm{H} := \mathrm{G}$. Let Σ be a covariance function on $[0,1]^2$ and $y \in \mathbf{Y}$. Each kernel $\mathbf{q}_{\mathrm{I}}$ is a Gaussian distribution on $\mathbf{T}_{\mathrm{I}}$ with mean vector $y_{\mathrm{I}} = \mathrm{H}_{\mathrm{I}} y$

and a covariance matrix defined by Σ. Then $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ is a conditional promeasure; whether it is tight on $\mathbf{T}$ depends on the choice of Σ. If the family is tight, $\mathbf{q} = \mathrm{G}^{\#} \varprojlim \langle \mathbf{q} \rangle_{\Gamma}$ is a family of Gaussian processes. More precisely, $\mathbf{q}(\,.\,, y)$ is a Gaussian process with paths $L_2[0,1]$, expected function y and covariance function Σ. If $\mathbf{p}_{\mathrm{I}}$ is chosen as a suitable Gaussian likelihood, each posterior $\widehat{\mathbf{q}}_{\mathrm{I}}$ is a location family of Gaussians, and by (2.7), $\widehat{\mathbf{q}}$ is a family of Gaussian processes. See Sec. 3.4 for details.

2.3. *Conjugate posteriors.* A Bayesian model is *conjugate* if the posterior belongs to the same class of distributions as the prior [45]. The limit in Theorem 2.5 preserves conjugacy—part (i) shows that conjugacy of all $(\mathbf{p}_{\mathrm{I}}, \mathbf{q}_{\mathrm{I}})$ implies conjugacy of $(\mathbf{p}, \mathbf{q})$, as in Example 2.6. There is, however, a stronger form of conjugacy which is of greater practical importance, since it permits the computation of posteriors even when Bayes' theorem is not applicable: A hyperparameter specifying the posterior can be computed as a function of observed data and of a prior hyperparameter. Lijoi and Prünster [37] call such models *parametrically conjugate*, since they are characterized by updates of a (hyper-)parameter. Virtually all commonly used Bayesian nonparametric models are parametrically conjugate [49]. An immediate question is hence whether the limit also preserves parametric conjugacy. The next result addresses this question and the closely related concept of sufficiency.

Given a measurable mapping S from $\mathbf{X}$ into a Polish space $\mathbf{S}$ and a probability kernel $\mathbf{v} : \mathbf{S} \to \mathbf{M}(\mathbf{X})$, define the set of measures

$$M_{\mathrm{S},\mathbf{v}} := \{\mu \in \mathbf{M}(\mathbf{X}) \,|\, \mu[\,.\,|S = s] =_{\text{a.s.}} \mathbf{v}(\,.\,, s)\} \,. \tag{2.8}$$

Recall that S is called a *sufficient statistic* for a set $M_0 \subset \mathbf{M}(\mathbf{X})$ if $M_0 \subset M_{\mathrm{S},\mathbf{v}}$ for some kernel $\mathbf{v}$. We call $\mathbf{v}$ a *sufficient kernel*. The statistic is sufficient for a kernel $\mathbf{p} : \mathbf{T} \to \mathbf{M}(\mathbf{X})$ if the image $\mathbf{p}(\,.\,, \mathbf{T})$ is contained in $M_{\mathrm{S},\mathbf{v}}$.

The next result involves the notion of a projective family of mappings: Let $\mathbf{S}$ be a Polish space embedded into a projective system $\langle \mathbf{S}_{\mathrm{I}}, e_{\mathrm{JI}} \rangle_{\Gamma}$ by mappings E_{I}. Suppose the measurable mappings $S_{\mathrm{I}} : \mathbf{X}_{\mathrm{I}} \to \mathbf{S}_{\mathrm{I}}$ are projective, i.e. $S_{\mathrm{I}} \circ f_{\mathrm{JI}} = e_{\mathrm{JI}} \circ S_{\mathrm{J}}$ for $I \preceq J$. Then there is a unique, measurable projective limit mapping $S_{\Gamma} : \mathbf{X}_{\Gamma} \to \mathbf{S}_{\Gamma}$. The mapping S_{Γ} has a well-defined *pull-back* $(\mathrm{F}, \mathrm{E})^{\#} S_{\Gamma} := \mathrm{E}^{-1} \circ S_{\Gamma} \circ \mathrm{F}$ if the image of $S_{\Gamma} \circ \mathrm{F}$ is contained in $\mathrm{E}\mathbf{S}$. Hence, the family $\langle S_{\mathrm{I}} \rangle_{\Gamma}$ determines a unique limit mapping $S : \mathbf{X} \to \mathbf{S}$, given by

$$S = (\mathrm{F}, \mathrm{E})^{\#} \varprojlim \langle S_{\mathrm{I}} \rangle_{\Gamma} \,. \tag{2.9}$$

Since $\mathbf{S}$ is Polish, the pull-back mapping S is measurable [30, Corollary 15.2].

Bayesian inference can be represented abstractly as a mapping from prior measure and observations to the posterior: We call a measurable mapping

$T\colon \mathbf{X}\times\mathbf{Y}\to\widehat{\mathbf{Y}}$, where $\widehat{\mathbf{Y}}$ is a Polish space, a *posterior index* if there exists a probability kernel $\mathbf{u}:\widehat{\mathbf{Y}}\to\mathbf{M}(\mathbf{T})$ such that

$$\widehat{\mathbf{q}}(\,.\,,x,y) =_{\text{a.s.}} \mathbf{u}(\,.\,,T(x,y))\;. \tag{2.10}$$

Clearly, $\mathrm{Id}_{\mathbf{X}\times\mathbf{Y}}$ is a trivial posterior index for any model. The model is called *parametrically conjugate* if there is a posterior index T for which

$$\widehat{\mathbf{Y}}\subset\mathbf{Y}\qquad\text{and}\qquad \mathbf{u}=_{\text{a.s.}}\mathbf{q}\;. \tag{2.11}$$

In the next result, $\mathbf{Y}$ is again embedded into $\langle\mathbf{Y}_{\mathrm{I}},h_{\mathrm{JI}}\rangle_{\Gamma}$ by $\langle\mathrm{H}_{\mathrm{I}}\rangle_{\Gamma}$, and the kernels $\mathbf{q}_{\mathrm{I}}$ are of the form $\mathbf{Y}_{\mathrm{I}}\to\mathbf{M}(\mathbf{T}_{\mathrm{I}})$.

THEOREM 2.7. *Require that the spaces $\mathbf{X}$, $\mathbf{T}$, $\mathbf{Y}$ and $\mathbf{S}$ are Polish. Let $\langle\mathbf{p}_{\mathrm{I}}\circ\mathrm{G}_{\mathrm{I}}\rangle_{\Gamma}$ and $\langle\mathbf{q}_{\mathrm{I}}\circ\mathrm{H}_{\mathrm{I}}\rangle_{\Gamma}$ be conditional promeasures which are tight on $\mathbf{X}$ and $\mathbf{T}$, respectively, and denote their limits by $\mathbf{p}$ and $\mathbf{q}$.*

(i) Let $\langle S_{\mathrm{I}}\rangle_{\Gamma}$ be a family of measurable mappings $\mathbf{X}_{\mathrm{I}}\to\mathbf{S}_{\mathrm{I}}$ with limit $S:\mathbf{X}\to\mathbf{S}$ in the sense of (2.9), and let $\langle\mathbf{v}_{\mathrm{I}}\rangle_{\Gamma}$ be a conditional promeasure of kernels $\mathbf{v}_{\mathrm{I}}\colon\mathbf{S}_{\mathrm{I}}\to\mathbf{M}(\mathbf{X}_{\mathrm{I}})$. Then

$$\mathbf{p}_{\mathrm{I}}(\,.\,,\mathbf{T}_{\mathrm{I}})\subset M_{\mathrm{S}_{\mathrm{I}},\mathbf{v}_{\mathrm{I}}}\quad\text{for all } I\in\Gamma\qquad\Longleftrightarrow\qquad\mathbf{p}(\,.\,,\mathbf{T})\subset M_{\mathrm{S},\mathbf{v}}\;, \tag{2.12}$$

for $\mathbf{v}:=\mathrm{F}^{\#}\varprojlim\langle S_{\mathrm{I}}\rangle_{\Gamma}$. The left-hand side implies $\langle\mathbf{v}_{\mathrm{I}}\rangle_{\Gamma}$ is tight on $\mathbf{X}$.

(ii) Let $\langle\widehat{\mathbf{q}}_{\mathrm{I}}\rangle_{\Gamma}$ be a conditional promeasure. A mapping T is a posterior index of $(\mathbf{p},\mathbf{q})$ if there is a posterior index T_{I} for each $(\mathbf{p}_{\mathrm{I}},\mathbf{q}_{\mathrm{I}})$ with

$$T=(\mathrm{F}\otimes\mathrm{H},\mathrm{H})^{\#}\varprojlim\langle T_{\mathrm{I}}\rangle_{\Gamma} \tag{2.13}$$

Conversely, if T is a posterior index for $(\mathbf{p},\mathbf{q})$ and if $\widehat{\mathbf{Y}}$ is embedded into a projective system $\langle\widehat{\mathbf{Y}}_{\mathrm{I}},\widehat{h}_{\mathrm{JI}}\rangle_{\Gamma}$ by mappings $\langle\widehat{\mathrm{H}}_{\mathrm{I}}\rangle_{\Gamma}$, there are posterior indices T_{I} for each induced model $(\mathbf{p}_{\mathrm{I}}=\mathrm{F}_{\mathrm{I}\#}\mathbf{p},\mathbf{q}_{\mathrm{I}}=\mathrm{G}_{\mathrm{I}\#}\mathbf{q})$ which satisfy (2.13). In particular, the model $(\mathbf{p},\mathbf{q})$ is parametrically conjugate if and only if each model $(\mathbf{p}_{\mathrm{I}},\mathbf{q}_{\mathrm{I}})$ is parametrically conjugate.

Thus, parametric conjugacy of a nonparametric Bayesian model can be guaranteed by an appropriate choice of conjugate marginal models $(\mathbf{p}_{\mathrm{I}},\mathbf{q}_{\mathrm{I}})$.

2.4. *Exponential family marginals.* In the parametric case, the most important class of conjugate Bayesian models are those defined by exponential family models and their canonical conjugate priors [45, Chapter 2]. For these models, the conditional densities of $\mathbf{p}_{\mathrm{I}}$ and $\mathbf{q}_{\mathrm{I}}$ with respect to suitable carrier measures are given by

$$p_{\mathrm{I}}(x_{\mathrm{I}}|\theta_{\mathrm{I}})=\frac{e^{\langle S_{\mathrm{I}}(x_{\mathrm{I}}),\theta_{\mathrm{I}}\rangle}}{Z_{\mathrm{I}}(\theta_{\mathrm{I}})}\qquad\text{and}\qquad q_{\mathrm{I}}(\theta_{\mathrm{I}}|\lambda,y_{\mathrm{I}})=\frac{e^{\langle\theta_{\mathrm{I}},\gamma_{\mathrm{I}}\rangle-\lambda\log Z_{\mathrm{I}}(\theta_{\mathrm{I}})}}{Z'_{\mathrm{I}}(\lambda,\gamma_{\mathrm{I}})}\;, \tag{2.14}$$

where Z_{I} and Z'_{I} are normalization functions. Each function $S_{\mathrm{I}} : \mathbf{X}_{\mathrm{I}} \to \mathbf{S}_{\mathrm{I}}$ is a sufficient statistic for $\mathbf{p}_{\mathrm{I}}$. Its range is a Polish vector space $\mathbf{S}_{\mathrm{I}}$ with inner product $\langle\, .\, ,\, .\,\rangle$ and contains the parameter space $\mathbf{T}_{\mathrm{I}}$ as a subspace. The prior is parameterized by a concentration $\lambda \in \mathbb{R}_+$ and by an expectation γ_{I} in the convex hull of $S_{\mathrm{I}}(\mathbf{X}_{\mathrm{I}})$.

If $\mathbf{p}_{\mathrm{I}}$ is given by (2.14) and the dimension of $\mathbf{T}_{\mathrm{I}}$ increases along sequences $I_1 \preceq I_2 \preceq \dots$, the limit model is nonparametric with infinite-dimensional parameter space $\mathbf{T}$. With Theorem 2.7, we obtain the nonparametric posterior, which is reminiscent of the characterization of conjugate parametric posteriors due to Diaconis and Ylvisaker [21, Chapter 1.3]:

COROLLARY 2.8. *If $\mathbf{p}_{\mathrm{I}}$ and $\mathbf{q}_{\mathrm{I}}$ in Theorem 2.7 are of the form* (2.14)*, and if the family $\langle S_{\mathrm{I}}\rangle_{\Gamma}$ admits a limit $S = (\mathrm{F},\mathrm{E})^{\#} \varprojlim \langle S_{\mathrm{I}}\rangle_{\Gamma}$ in the sense of* (2.9)*, the model $(\mathbf{p},\mathbf{q})$ is parametrically conjugate with*

$$T(X,\lambda,\gamma) = \Big(\lambda + 1, \gamma + S(X)\Big)\ . \tag{2.15}$$

A specific example of (2.15) is the Dirichlet process with concentration α and base measure μ, for which T is given by

$$(x^{(1)},\dots,x^{(n)},\alpha\cdot\mu) \ \mapsto\ \frac{1}{n+\alpha}\Big(\alpha\mu + \sum_{k=1}^{n}\delta_{x^{(k)}}\Big)\ . \tag{2.16}$$

Due to the parameterization of the Dirichlet distribution, the choice of λ does not affect the form of the model, and $\gamma = \alpha\mu$. Gaussian process models, and in fact most models in the nonparametric Bayesian literature, can also be interpreted in this manner. Section 3 provides further illustration.

2.5. *Ergodic decomposition and conditional independence.* Suppose observational data is expressed by random variables X^y, each with law $\mathbf{r}(\,.\,,y)$ governed by some configuration $y \in \mathbf{Y}$. From an abstract point of view, a Bayesian model that accounts for the data is a family of random measures $\Pi^y : \Omega \to \mathbf{M}(\mathbf{X})$ satisfying

$$\mathbf{r}(\,.\,,y) = \int_{\Omega} \Pi^y(\omega)\mathbb{P}(d\omega) = \mathbb{E}[\Pi^y]\ . \tag{2.17}$$

To construct such a model, we have to identify a suitable parameter space $\mathbf{T}$ and a decomposition $\Pi^y = \mathbf{p}\circ\Theta^y$. If the variables X^y are exchangeable random sequences, de Finetti's theorem can be invoked for guidance, but not if the observations aggregate into another type of structure—e.g. a large graph or permutation, as in Sec. 3. This problem naturally leads to ergodic decompositions and to the work of Lauritzen [36] on extremal families.

Suppose $S : \mathbf{X} \to \mathbf{S}$ is a statistic with values in a Polish space, and define the image measures $\rho^y := S_\#\mathbf{r}(\,.\,, y)$. Standard results guarantee each measure $\mathbf{r}(\,.\,, y)$ has a (S, ρ^y)-disintegration, i.e. there is a probability kernel $\nu^y : \mathbf{S} \to \mathbf{M}(\mathbf{X})$ such that $\mathbf{r}(\,.\,, y) = \int_{\mathbf{S}} \nu^y(\,.\,, s)\rho^y(ds)$. Thus, the random measures defined by $\tilde{\Pi}^y := \nu^y \circ (S \circ X^y)$ satisfy (2.17), but the result is trivial: The kernels ν^y depend on y. Theorem 2.9 below shows that a much stronger result holds if the statistic is sufficient and has ergodic measures: In this case, ρ^y concentrates on a subset $\mathbf{T}$ of $\mathbf{S}$, independent of y, and all disintegrations ν^y coincide.

Let $\mathbf{v} : \mathbf{S} \to \mathbf{M}(\mathbf{X})$ be a kernel and $M_{\mathrm{S},\mathbf{v}}$ defined as in (2.8). We will find that this set is convex and denote its set of extreme points by $\mathrm{ex}(M_{\mathrm{S},\mathbf{v}})$. Let $\mathcal{S} := S^{-1}\mathcal{B}(\mathbf{S})$ be the sufficient σ-algebra in $\mathcal{B}(\mathbf{X})$ and

$$\mathcal{E}_{\mathrm{S},\mathbf{v}} := \{\eta \in M_{\mathrm{S},\mathbf{v}} \,|\, \forall A \in \mathcal{S} :\ \eta(A) \in \{0, 1\}\} \,. \tag{2.18}$$

Measures in $\mathcal{E}_{\mathrm{S},\mathbf{v}}$ are called *S-ergodic*. Let further $\Delta_{\mathrm{S},\mathbf{v}}$ be the set of all distributions in $M_{\mathrm{S},\mathbf{v}}$ under which the distribution of S is degenerate,

$$\Delta_{\mathrm{S},\mathbf{v}} := \{\eta \in M_{\mathrm{S},\mathbf{v}} \,|\, \exists s_\eta \in \mathbf{S} :\ S_\#\eta = \delta_{s_\eta}\} \,. \tag{2.19}$$

The definition induces a mapping $\tau : \Delta_{\mathrm{S},\mathbf{v}} \to \mathbf{S}$ given by $\eta \mapsto s_\eta$. Its image is denoted $\mathbf{T} := \tau(\Delta_{\mathrm{S},\mathbf{v}})$. The next result is an ergodic decomposition in the spirit of [36, 15], adapted to be applicable to nonparametric Bayesian models defined on embedded spaces. It shows how sufficient statistics can be used to identify a suitable decomposition of the model.

THEOREM 2.9. *Let $\mathbf{X}$ and $\mathbf{S}$ be Polish and S a sufficient statistic for $\mathbf{r}$, with sufficient kernel $\mathbf{v}$. Then $M_{\mathrm{S},\mathbf{v}}$ is convex and*

$$ex(M_{\mathrm{S},\mathbf{v}}) = \mathcal{E}_{\mathrm{S},\mathbf{v}} = \Delta_{\mathrm{S},\mathbf{v}} \,. \tag{2.20}$$

If $M_{\mathrm{S},\mathbf{v}}$ is weakly closed, the set $\mathbf{T} := \tau(\Delta_{\mathrm{S},\mathbf{v}})$ is a standard Borel space. There is a unique kernel $\mathbf{q} : \mathbf{Y} \to \mathbf{M}(\mathbf{T})$ such that the restriction $\mathbf{p} := \mathbf{v}|_{\mathbf{T}}$ is a $(S, \mathbf{q}(\,.\,, y))$-disintegration of $\mathbf{r}(\,.\,, y)$ for all $y \in \mathbf{Y}$. The map τ is a parameterization of $\mathcal{E}_{\mathrm{S},\mathbf{v}}$, i.e. a Borel isomorphism of $\mathcal{E}_{\mathrm{S},\mathbf{v}}$ and $\mathbf{T}$.

The "Bayesian" interpretation of Theorem 2.9 is that $\mathbf{T} = \tau(\mathcal{E}_{\mathrm{S},\mathbf{v}})$ is a parameter space, $\mathbf{Y}$ is a set of hyperparameters, and there is a uniquely defined family of priors $\mathbf{q}(\,.\,, y)$ such that

$$\mathbf{r}(\,.\,, y) =_{\text{a.s.}} \int_{\mathbf{T}} \mathbf{p}(\,.\,, \theta)\mathbf{q}(d\theta, y) \,. \tag{2.21}$$

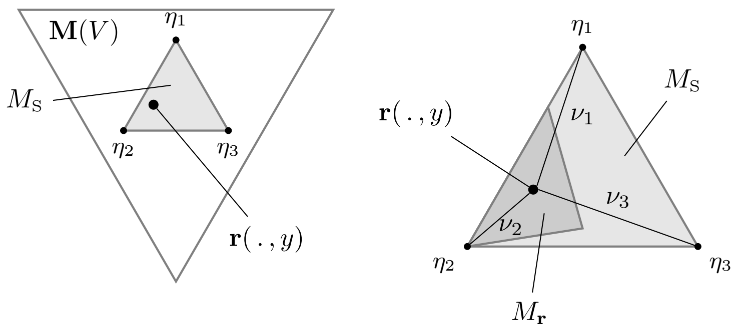

FIG 1. *Finite analogue of Theorem 2.9 for a probability space with three disjoint events. Left: The space $\mathbf{M}(V)$ of probability measures contains $M_{\mathrm{S},\mathbf{v}}$ as a convex subset. The ergodic measures are $\eta_j = \mathbf{v}(\,.\,,\theta = j)$. Right: Each measure $\mathbf{p}(\,.\,,y)$ is a convex combination with coefficients $\nu_j = \mathbf{q}(\{\eta_j\},y)$. By Theorem 2.7, the set $M_{\mathbf{r}} := \{\mathbf{r}(\,.\,,y)|y \in \mathbf{Y}\}$ is contained in $M_{\mathrm{S},\mathbf{v}}$. Even if $M_{\mathbf{r}} \subsetneq M_{\mathrm{S},\mathbf{v}}$, $\mathbf{q}(\,.\,,y)$ may assign non-zero mass to ergodic measures $\eta \notin M_{\mathbf{r}}$.*

The choice of $\mathbf{S}$ determines by what type of mathematical objects—functions, measures, etc—the model is parameterized. Since $\mathcal{E}_{\mathrm{S},\mathbf{v}} \subset M_{\mathrm{S},\mathbf{v}}$, the family $\mathcal{E}_{\mathrm{S},\mathbf{v}}$ (and hence the kernel $\mathbf{p}$) inherits S as a sufficient statistic from $M_{\mathrm{S},\mathbf{v}}$.

The integral decomposition (2.21) does not quite imply that $M_{\mathrm{S},\mathbf{v}}$ is a Choquet simplex, since the set need not be compact. Rather, existence of $\mathbf{q}$ in Theorem 2.9 follows from a non-compact generalization of Choqet's theorem, but uniqueness holds only given the specific choice of $\mathbf{v}$.

2.6. *Extremal families in Bayesian nonparametrics.* We have not yet considered how to define the sufficient statistic S and the sufficient kernel $\mathbf{v}$ in Theorem 2.9. Lauritzen [36] proposed to obtain $\mathbf{v}$ as a projective limit, and S as an associated limit of sufficient statistics. His ideas are of direct relevance to the construction of nonparametric priors.

REMARK 2.10. Specifically, Lauritzen [36] addresses the following case: Require $\mathbf{X} = \mathbf{X}_{\Gamma}$. Given is a set $M \subset \mathbf{M}(\mathbf{X}_{\Gamma})$. Suppose there is a continuous statistic $S_{\mathrm{I}} : \mathbf{X}_{\mathrm{I}} \to \mathbf{S}_{\mathrm{I}}$ for each $I \in \Gamma$, such that (i) each S_{I} is sufficient for the set $f_{\mathrm{I}\#}M \subset \mathbf{M}(\mathbf{X}_{\mathrm{I}})$, (ii) each of the associated sufficient kernels $\mathbf{v}_{\mathrm{I}} : \mathbf{S}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ is continuous, and (iii) the family $\langle \mathbf{v}_{\mathrm{I}} \rangle_{\Gamma}$ is projective. Then the tail σ-algebra $\mathcal{S}_{\mathrm{T}} := \limsup_{\mathrm{I}\in\Gamma} \sigma(S_{\mathrm{I}})$ is sufficient for M. The perhaps more widely known work of Diaconis and Freedman [14, 15] addresses the sequential special case $\mathbf{X}_{\Gamma} = \mathbf{X}_0^{\infty}$, where $\mathbf{X}_0$ is Polish. In Lauritzen's nomenclature, a conditional promeasure of continuous kernels is called a *projective statistical field*, and the set $\mathcal{E}$ of ergodic measures an *extremal family*.

The approach requires only minor modifications for application to Bayesian nonparametrics: Theorems 2.3 and 2.9 permit us to discard the continuity assumptions and the requirement $\mathbf{X} = \mathbf{X}_\Gamma$. Suppose $\mathbf{r}$ is given as $\mathbf{r} = \mathrm{F}^{\#} \varprojlim \langle \mathbf{r}_{\mathrm{I}} \rangle_\Gamma$ and there is a sufficient statistic S_{I} for each $\mathbf{r}_{\mathrm{I}}$. If the sufficient kernels $\mathbf{v}_{\mathrm{I}}$ form a tight conditional promeasure $\langle \mathbf{v}_{\mathrm{I}} \rangle_\Gamma$ on $\mathbf{X}$, the limit kernel $\mathbf{v}$ is by Theorem 2.3 a conditional probability given the tail σ-algebra $\mathcal{S}_{\mathrm{T}} := \limsup_{\mathrm{I}\in\Gamma} \sigma(S_{\mathrm{I}})$. If the limit statistic S generates $\mathcal{S}_{\mathrm{T}}$, it is hence a sufficient statistic for $\mathbf{r}$, and the pair $(S, \mathbf{v})$ can be substituted into Theorem 2.9. We focus on a specific form of limit statistic:

Corollary 2.11. *Let $\langle \mathbf{r}_{\mathrm{I}} \rangle_\Gamma$ be a tight conditional promeasure on $\mathbf{X}$. Let $\mathbf{S}$ be a Polish space and $\langle S_{\mathrm{I}} : \mathbf{X}_{\mathrm{I}} \to \mathbf{S} \rangle_\Gamma$ a family of measurable mappings which converge point-wise along a cofinal sequence $I_1 \preceq I_2 \preceq \dots$ in Γ,*

$$S(x) = \lim_{n\to\infty} S_{\mathrm{I_n}}(\mathrm{F}_{\mathrm{I_n}} x) \quad \text{for all } x \in \mathbf{X}\,. \tag{2.22}$$

If each S_{I} is sufficient for $\mathbf{r}_{\mathrm{I}}$, and if the sufficient kernels $\mathbf{v}_{\mathrm{I}} : \mathbf{S} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ form a tight conditional promeasure on $\mathbf{X}$, then the limits S, $\mathbf{r} = \mathrm{F}^{\#} \varprojlim \langle \mathbf{r}_{\mathrm{I}} \rangle_\Gamma$ and $\mathbf{v} = \mathrm{F}^{\#} \varprojlim \langle \mathbf{v}_{\mathrm{I}} \rangle_\Gamma$ satisfy Theorem 2.9.

De Finetti's theorem [28, Theorem 11.10] provides an illustrative example: Choose $\mathbf{X} = \prod_{n=1}^{\infty} \mathbf{X}_0$ and $S(x)$ as the empirical measure of a sequence $x \in \mathbf{X}$. The latter implies permutation invariance of random sequences. Thus, $M_{\mathrm{S},\mathbf{v}}$ contains the exchangeable distributions, $\mathcal{E}_{\mathrm{S},\mathbf{v}}$ the factorial measures, (2.21) is the de Finetti mixture representation and (2.20) is the Hewitt-Savage zero-one law [29, Theorem 3.15].

If the data does not have the structure of an exchangeable sequence, another sufficient statistic can be substituted for the empirical measure to determine conditional independence in form of the decomposition (2.21). The limit statistic permits the ergodic measures to be identified conveniently as the degenerate measures $\Delta_{\mathrm{S},\mathbf{v}}$. See Corollaries 3.8 and 3.5 for examples.

Although we cannot generally assume continuity of the kernels, Lauritzen's continuity assumption does, where applicable, simplify Theorem 2.9:

Proposition 2.12. *Suppose $\mathbf{v} = \mathrm{F}^{\#} \langle \mathbf{v}_{\mathrm{I}} \rangle_\Gamma$ in Theorem 2.9. If all mappings $\mathbf{v}_{\mathrm{I}} : \mathbf{S}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ and the limit statistic S are continuous, $M_{\mathrm{S},\mathbf{v}}$ is a weakly closed set.*

The following *parametric* example of an ergodic decomposition is due to [35] and [16]. Suppose Γ consists of the sets $I = [n]$, where $n \in \mathbb{N}$ denotes sample size. Let $\mathbf{X}_{\mathrm{I}}$ be product space $\mathbf{X}_{\mathrm{I}} = (\mathbb{R}^d)^n$ and let $\mathbf{S}_{\mathrm{I}} = \mathbf{S} = \mathbb{R}^k$

for some fixed, finite dimensions $d, k \in \mathbb{N}$. Since $\mathbf{X} = (\mathbb{R}^d)^\infty$ coincides with the projective limit space $\mathbf{X}_\Gamma$, tightness conditions are not required. Under suitable regularity conditions, exponential families can be obtained ergodic measures in Theorem 2.9 by defining S_{I} as

$$S_{[n]}(x_1, \ldots, x_n) := \sum_{i \in [n]} S_0(x_i) \tag{2.23}$$

for a function $S_0 : \mathbf{X}_0 \to \mathbb{R}^k$ [35, 16]. The conditional density of the kernel $\mathbf{p}$ in (2.21) is then of the form (2.14), where $\theta_{[n]} = \theta \in \mathbb{R}^k$. See also [15].

Thus, exponential families can arise in two distinct ways in our results: (1) As finite-dimensional marginals in Theorem 2.7, in which case the dimension of $\mathbf{S}_{\mathrm{I}}$ (and hence of $\mathbf{T}_{\mathrm{I}}$) increases with sample size. Hence, the limit model is nonparametric. (2) As ergodic measures in Theorem 2.9. Here, $\mathbf{S} = \mathbf{S}_{\mathrm{I}}$ has constant finite dimension, and the limit model is parametric.

3. Applications and examples. We apply our results to several problems of interest to Bayesian nonparametrics: Random measures, which illustrate the construction by a well-known problem; random permutations and virtual permutations; random infinite graphs; and the widely used Gaussian process regression model on random functions.

3.1. *Random Probability Measures.* Let V be a Polish space with Borel sets $\mathcal{B}_V$. Our objective is to construct a family of random measures Θ^y, each with distribution $\mathbf{q}(\,.\,, y)$, as frequently used as nonparametric priors. Let $\mathbf{T} = \mathbf{M}(V)$. A suitable embedding into a projective system $\langle \mathbf{T}_{\mathrm{I}}, g_{\mathrm{JI}} \rangle_\Gamma$ by means of mappings G_{I} be constructed as follows [41]: Fix a countable dense subset $V' \subset V$ and a metric d compatible with the topology on V. Let $\mathcal{Q} \subset \mathcal{B}_V$ be the countable algebra generated by the open d-balls with rational radii and centers in V. This algebra is a countable generator of $\mathcal{B}_V$. Let Γ be the set of all partitions of V which consist of a finite number of sets in $\mathcal{Q}$. For any two partitions $I, J \in \Gamma$, define $I \preceq J$ if and only if $I \cap J = J$. Then $(\Gamma, \preceq)$ is a partially ordered, directed, countable set. For each partition $I = (A_1, \ldots, A_n)$ in Γ, let $\mathbf{T}_{\mathrm{I}} := \mathbf{M}(I)$ be the set of all probability measures on the finite σ-algebra generated by the sets in partition I, that is, $\mathbf{T}_{\mathrm{I}}$ is the unit simplex in $\mathbb{R}^{\mathrm{I}}$.

Choose $\mathrm{G}_{\mathrm{I}} : \mathbf{M}(V) \to \mathbf{M}(I)$ as the evaluation $\mu \mapsto (\mu(A_1), \ldots, \mu(A_n))$. The σ-algebra $\sigma(\mathrm{G}_{\mathrm{I}}; I \in \Gamma)$ which the evaluations generate on $\mathbf{M}(V)$ coincides with the Borel σ-algebra of the weak* topology on $\mathbf{M}(V)$. For each pair $I \preceq J$ of partitions, define g_{JI} as the unique mapping satisfying $\mathrm{G}_{\mathrm{I}} = g_{\mathrm{JI}} \circ \mathrm{G}_{\mathrm{J}}$. Then $\langle \mathbf{T}_{\mathrm{I}}, g_{\mathrm{JI}} \rangle_\Gamma$ is a projective system and $\mathbf{T}_\Gamma = \varprojlim \langle \mathbf{T}_{\mathrm{I}} \rangle_\Gamma$ is the set of probability charges (finitely additive probabilities) on $\mathcal{Q}$. The projective limit

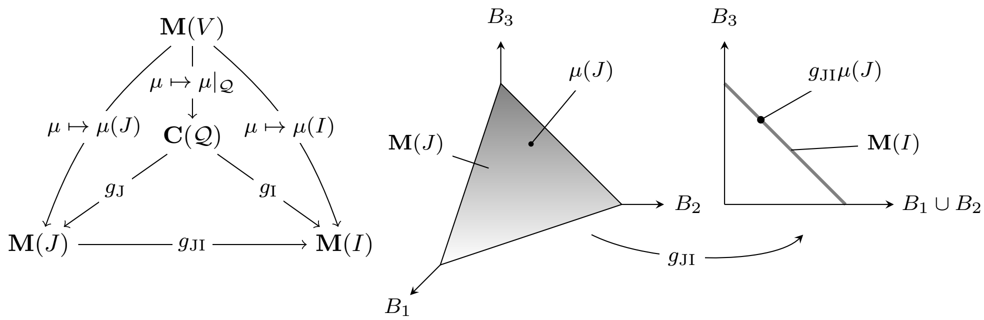

FIG 2. *Left: Spaces in the random measure construction.* $\mathbf{M}(V)$ *denotes the space of probability measures on* V*,* $\mathbf{C}(\mathcal{Q})$ *the set of probability charges on the countable algebra* $\mathcal{Q} \subset \mathcal{B}(V)$*, and* $\triangle_{\mathrm{I}}$ *the standard simplex in* $\mathbb{R}^{\mathrm{I}}$*. Middle: The simplex* $\triangle_{\mathrm{J}}$ *for a partition* $J = (B_1, B_2, B_3)$*. Right: A new simplex* $\triangle_{\mathrm{I}} = g_{\mathrm{JI}}\triangle_{\mathrm{J}}$ *is obtained by merging the sets* B_1 *and* B_2 *to produce* $I = (B_1 \cup B_2, B_3)$*.*

mapping $\mathrm{G} = \varprojlim \langle \mathrm{G}_{\mathrm{I}} \rangle_{\Gamma}$ restricts measures on $\mathcal{B}_V$ to the subsystem $\mathcal{Q}$, and embeds $\mathbf{M}(V)$ into $\mathbf{T}_{\Gamma}$ as a measurable subset [41, Proposition 3.1].

To construct a conditional promeasure, choose a hyperparameter space $\mathbf{Y}$ and define kernels $\mathbf{q}_{\mathrm{I}} : \mathbf{Y}_{\mathrm{I}} \to \mathbf{M}(\mathbf{T}_{\mathrm{I}})$ such that they satisfy the following two conditions: Require that, for all $y \in \mathbf{Y}$, there exists $\mu^y \in \mathbf{M}(V)$ such that

$$g_{\mathrm{JI}\#}\mathbf{q}_{\mathrm{J}}(\,.\,, \mathrm{H}_{\mathrm{J}}y) =_{\text{a.s.}} \mathbf{q}_{\mathrm{I}}(\,.\,, \mathrm{H}_{\mathrm{I}}y) \quad \text{and} \quad \int_{\mathbf{T}_{\mathrm{I}}} \theta_{\mathrm{I}}\mathbf{q}_{\mathrm{I}}(d\theta_{\mathrm{I}}, y) = \mathrm{F}_{\mathrm{I}\#}\mu^y\,. \tag{3.1}$$

The first condition makes $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ a conditional promeasure. The second condition ensures that the projective limit $\mathbf{q}_{\Gamma}$ assigns outer measure 1 to the subset of those charges in $\mathbf{T}_{\Gamma}$ which are countable additive [41, Proposition 4.1]. Hence, by Lemma A.2, $\mathbf{q}$ is tight on $\mathbf{M}(V)$. A consequence of the second condition is that the constructed random measures have expectations $\mathbb{E}[\Theta^y] = \mu^y$. To ensure the condition holds, $\mathbf{q}_{\mathrm{I}}(\,.\,, y)$ can be parameterized in terms of μ^y, as the example of the Dirichlet process shows.

EXAMPLE 3.1 (Dirichlet process). Let $\mathbf{Y} := \mathbb{R}_+ \times \mathbf{M}(V)$ and define $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ as follows: For any $(\alpha, \mu) \in \mathbf{Y}$, choose $\mathbf{q}_{\mathrm{I}}(\,.\,, (\alpha, \mu))$ as the Dirichlet distribution on the simplex $\mathbf{T}_{\mathrm{I}} \subset \mathbb{R}^{\mathrm{I}}$, with concentration α and expectation $\mathrm{F}_{\mathrm{I}\#}\mu$. Then $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ satisfies the conditions (3.1). Let $\mathbf{q} = \mathrm{G}^{\#} \varprojlim \langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$. For any given α and μ, the measure $\mathbf{q}(\,.\,, \alpha\mu)$ is the Dirichlet process of Ferguson [19], with concentration α and "base measure" μ. By Theorem 2.7, the posterior is updated under observations according to (2.16).

3.2. *Random Permutations.* A prominent infinite random permutation is the *Chinese Restaurant Process* (CRP), defined by Dubins and Pitman

[43] as the law given by uniform marginals on the finite symmetric groups $\mathbb{S}_n$. Its distribution induces a distribution on partitions which is now a mainstay of Bayesian nonparametric statistics. More recently, infinite random permutations have been considered heuristically in the computer science literature [40] for nonparametric approaches to preference data analysis: Preference lists are represented as permutations, and a nonparametric approach requires a coherent extension to an infinite number of items or choices. This problem motivates Example 3.2 below.

As a common setting for both these cases, we use a beautiful projective limit construction due to Kerov, Olshanski, and Vershik [33]. Denote by $\mathbb{S}_\infty$ the infinite symmetric group, i.e. the group of all permutations on $\mathbb{N}$ under which all but a finite number of elements remain invariant. Choose $\Gamma = \{[n] | n \in \mathbb{N}\}$, ordered by inclusion, and $\mathbf{X}_{[n]} := \mathbb{S}_n$. A projector $f_{[\mathrm{n+1}][\mathrm{n}]} : \mathbb{S}_{n+1} \to \mathbb{S}_n$ must consistently remove the entry $n+1$ from an element of $\mathbb{S}_{n+1}$, for instance, delete 4 from both rows of $\left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{smallmatrix}\right)$. Kerov *et al.* [33] define $f_{[\mathrm{n+1}][\mathrm{n}]}$ as deletion of $n+1$ from its cycle: Any permutation $\pi \in \mathbb{S}_n$ admits a unique representation of the form

$$\pi = \sigma_{k_1}(1)\sigma_{k_2}(2)\cdots\sigma_{k_n}(n) \ , \tag{3.2}$$

where k_i are natural numbers with $k_i \leq i$, and $\sigma_i(j)$ denotes the transposition of i and j. The map $\psi_n : \pi_{[n]} \mapsto (k_1, \ldots, k_n)$ is thus a bijection $\mathbb{S}_n \to \prod_{m \leq n} [m]$, and a homeomorphism of the discrete topologies on $\mathbb{S}_n$ and the product space. Hence, the pull-back $(\psi_{n+1}, \psi_n)^{\#} \mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]} =: f_{[\mathrm{n+1}][\mathrm{n}]}$ of the product space projector $\mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]}$ exists, and the diagram

$$\begin{tikzcd} \mathbb{S}_{n+1} \arrow[r, "\psi_{n+1}"] \arrow[d, "f_{[\mathrm{n+1}][\mathrm{n}]}"'] & \prod_{m \leq n+1} [m] \arrow[d, "\mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]}"] \\ \mathbb{S}_n \arrow[r, "\psi_n"] & \prod_{m \leq n} [m] \end{tikzcd} \tag{3.3}$$

commutes. Since application of $\sigma_{k_1}(1), \ldots, \sigma_{k_{n+1}}(n+1)$ from the left consecutively constructs the cycles of $\pi_{[n+1]}$, deletion of the final step by $\mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]}$ indeed amounts to removal of $n+1$ from its cycle. The definition of $f_{[\mathrm{n+1}][\mathrm{n}]}$ is consistent with the CRP: The image measure of the CRP marginal distribution on $\mathbb{S}_{n+1}$ under $f_{[\mathrm{n+1}][\mathrm{n}]}$ is the CRP marginal on $\mathbb{S}_n$.

Elements of the projective limit space $\mathfrak{S} := \mathbf{X}_\Gamma = \varprojlim \langle \mathbb{S}_n, f_{[\mathrm{n+1}][\mathrm{n}]} \rangle_\Gamma$ are infinite sequences $\pi = \sigma_{k_1}(1)\sigma_{k_2}(2)\cdots$. These are mappings which iteratively permute pairs of elements ad infinitum, and are called *virtual permutations* in [33]. The space $\mathfrak{S}$ compactifies $\mathbb{S}_\infty$: In the projective limit

$$\begin{tikzcd}
& \mathbb{S}_\infty \arrow[d, hook] \arrow[ddl, bend right, "f_{n+1}\big|_{\mathbb{S}_\infty}"'] \arrow[ddr, bend left, "f_n\big|_{\mathbb{S}_\infty}"] & \\
& \mathfrak{S} \arrow[dl, "f_{n+1}"'] \arrow[dr, "f_n"] & \\
\mathbb{S}_{n+1} \arrow[rr, "f_{[n+1][n]}"] & & \mathbb{S}_n
\end{tikzcd}$$

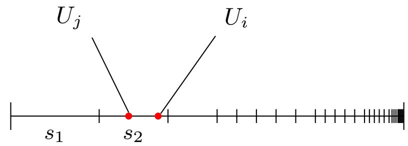

FIG 3. *Left: Embedding diagram for the permutation example. Since* $\mathbb{S}_\infty$ *is a subset of* $\mathfrak{S}$*, the mapping* F *is the canonical inclusion* $\mathbb{S}_\infty \hookrightarrow \mathfrak{S}$*. Right: Construction of a random virtual permutation* $\pi \sim \mathbf{p}(\,.\,, s)$ *given a mass partition* s*.*

topology, $\mathfrak{S}$ is a compact space and contains $\mathbb{S}_\infty$ as a dense subset [33, §1.2]. The embedding F in (2.2) is thus the canonical inclusion $\mathbb{S}_\infty \hookrightarrow \mathfrak{S}$. Fig. 3 shows the corresponding diagram. As a projective limit of homeomorphisms, $\psi := \varprojlim \langle \psi_{[n]} \rangle_\Gamma$ is a homeomorphism of $\mathfrak{S}$ and $\prod_{m\in\mathbb{N}}[m]$. If and only if $\pi \in \mathbb{S}_\infty$, the sequence of transpositions $\sigma_{k_1}(1)\sigma_{k_2}(2)\cdots$ becomes trivial after a finite number n of steps, and $\psi(\pi) = (k_1, \ldots, k_n, n+1, n+2, \ldots)$. The space $\mathfrak{S}$ is not a group: If $\psi(\pi) = (1, 1, 2, 3, \ldots)$, for example, π is not invertible.

EXAMPLE 3.2 (A nonparametric Bayesian model on $\mathbb{S}_\infty$). Let $\mathbf{T}_{[n]} = \mathbb{R}^n$. We define kernels $\mathbf{p}_{[n]} : \mathbb{R}^n \to \mathbf{M}(\mathbb{S}_n)$ as the conditional probabilities whose densities with respect to counting measure on $\mathbb{S}_n$ are

$$p_{[n]}(\pi_{[n]}|\theta_{[n]}) := \frac{1}{Z_{[n]}(\theta_{[n]})} \exp\Big(-\sum_{j=1}^{n} \theta_{[n]}^{(\mathrm{j})} W^{(\mathrm{j})}(\pi_{[n]})\Big) \qquad \theta_{[n]} \in \mathbb{R}^n\ , \tag{3.4}$$

where $W^{(\mathrm{j})}(\pi_{[n]}) := 1 - \{k_j = j\}$. Hence, $W^{(\mathrm{j})} = 0$ if and only if j is the smallest element on its cycle, and therefore $\sum_j W^{(\mathrm{j})} = n - \#\text{cycles}$. Clearly, $S_{[n]} := (-W^{(1)}, \ldots, -W^{(\mathrm{j})})$ is a sufficient statistic for (3.4). This parametric model was introduced as the *generalized Cayley model* by Fligner and Verducci [20]. To embed $\mathbf{T}$, define a projective system by choosing $g_{[\mathrm{n+1}][\mathrm{n}]} := \mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]}$, which implies $\mathbf{T}_\Gamma = \mathbb{R}^{\mathbb{N}}$.

The choice of $\mathbf{T}$ itself requires some further consideration to ensure tightness of $\langle \mathbf{p}_{[n]} \rangle_\Gamma$: If $\pi_{[n]}$ is distributed according to (3.4), the random variables $W^{(\mathrm{j})}(\pi_n)$ are independent [20, §2]. The partition function $Z_{[n]}$ thus factorizes as $Z_{[n]}(\theta) = \prod_j Z^{(\mathrm{j})}(\theta^{(\mathrm{j})}) = \prod_j M^{(\mathrm{j})}(-\theta^{(\mathrm{j})})$, where $M^{(\mathrm{j})}$ is the moment-generating function of the variable $W^{(\mathrm{j})}(\pi_{[n]})$. Hence, for each $j \in \mathbb{N}$,

$\mathrm{Law}(W^{(\mathrm{j})})$ depends only on $\theta^{(\mathrm{j})}$. We define

$$G_j(\theta^{(\mathrm{j})}) := \Pr\{W^{(\mathrm{j})}(\pi) = 1\} = \frac{(j-1)e^{-\theta^{(\mathrm{j})}}}{1+(j-1)e^{-\theta^{(\mathrm{j})}}} \tag{3.5}$$

for each j, and similarly $G(\theta) := (G_j(\theta^{(\mathrm{j})}))_{j\in\mathbb{N}}$ for sequences $\theta \in \mathbb{R}^{\mathbb{N}}$. Define hyperparameter spaces $\mathbf{Y}_{[n]} := \mathbb{R}_+ \times \mathbb{R}^n$ and a projective system $\langle \mathbf{Y}_{[n]}, h_{[\mathrm{n+1}][\mathrm{n}]} \rangle_\Gamma$ with $h_{[\mathrm{n+1}][\mathrm{n}]} := \mathrm{Id}_{\mathbb{R}_+} \otimes \mathrm{pr}_{[\mathrm{n+1}][\mathrm{n}]}$. Each prior $\mathbf{q}_{[n]}$ is now chosen as the natural conjugate prior of $\mathbf{p}_{[n]}$ with hyperparameter space $\mathbf{Y}_{[n]}$. Application of our results yields:

PROPOSITION 3.3. *The families $\langle \mathbf{p}_{[n]} \rangle_\Gamma$ and $\langle \mathbf{q}_{[n]} \rangle_\Gamma$ defined above are conditional promeasures, and $\langle \mathbf{p}_{[n]} \rangle_\Gamma$ is tight on $\mathbb{S}_\infty \subset \mathfrak{S}$ if and only if* $\mathrm{G}(\theta) \in \ell_1(0,1)$. *Similarly, $\langle \mathbf{q}_{[n]} \rangle_\Gamma$ is tight on $\mathbf{T} := \mathrm{G}^{-1}\ell_1(0,1)$ if and only if $\mathbf{Y} = \mathbb{R}_+ \times \mathrm{G}^{-1}\ell_1(0,1)$. In this case, the Bayesian model on $\mathbb{S}_\infty$ defined by $\mathbf{p} = \mathrm{F}^{\#} \varprojlim \langle \mathbf{p}_{[n]} \rangle_\Gamma$ and $\mathbf{q} = \mathrm{G}^{\#} \varprojlim \langle \mathbf{q}_{[n]} \rangle_\Gamma$ is conjugate with posterior index*

$$T(\pi, (\lambda, \gamma)) = \Big(\lambda + 1, \gamma + S(\pi)\Big) \tag{3.6}$$

and the posterior concentrates on $\mathbf{T}$.

The next example is an applies Theorem 2.9 to random permutations whose laws are invariant under relabeling of the permuted items. This is not a useful assumption for preference data, but rather concerns the relation between random permutations and random partitions: A virtual permutation $\pi \in \mathfrak{S}$ decomposes $\mathbb{N}$ into disjoint cycles $C_1, C_2, ...$ The corresponding unordered sets $B_1, B_2, \dots$ form a partition of $\mathbb{N}$, which we denote $\mathrm{part}(\pi)$. If and only if $\pi \in \mathbb{S}_\infty$, all blocks of $\mathrm{part}(\pi)$ are finite. Let $b_j = \lim_n \frac{1}{n}|B_j \cap [n]|$ denote the limiting relative block sizes, called the *asymptotic frequencies* [7]. For a partition Ψ of $\mathbb{N}$, let $|\,\Psi\,|^{\downarrow}$ be the vector of asymptotic frequencies, ordered by decreasing size. Then $|\,\Psi\,|^{\downarrow}$ is a sequence of non-negative numbers which sum to 1. Denote the set of such sequences by $\mathcal{P}_m$. Using the sequence $|\,\mathrm{part}(\pi)\,|^{\downarrow}$ as a sufficient statistic in Theorem 2.9 yields a "pull-back" of Kingman's paint-box theorem [7, Theorem 2.1] to virtual permutations:

EXAMPLE 3.4 (Ergodic decomposition of virtual permutations). Define a kernel $\mathbf{p} : \mathcal{P}_m \to \mathbf{M}(\mathfrak{S})$ as follows: A sequence $s = (s_1, s_2, \dots) \in \mathcal{P}_m$ partitions the unit interval into intervals $[\sum_{i=1}^{j} s_i, \sum_{i=1}^{j+1} s_i)$. A random draw π from $\mathbf{p}(\,.\,, s)$ is obtained by drawing $U_1, U_2, \dots \sim_{\mathrm{iid}} \mathrm{Uniform}[0,1]$. Then $i, j \in \mathbb{N}$ are in the same cycle of π if U_i, U_j are in the same interval of the partition defined by s. If so, i precedes j on the cycle if $U_i < U_j$ (Fig. 3). For $\sigma \in \mathbb{S}_\infty$, denote by λ_σ the conjugate action $\lambda_\sigma \pi := \sigma\pi\sigma^{-1}$ of σ on $\mathfrak{S}$.

COROLLARY 3.5. *Let Λ be the group of conjugations of $\mathbb{S}_\infty$ and $\mathbf{Y}$ a Polish space. Let $\mathbf{r} : \mathbf{Y} \to \mathbf{M}(\mathfrak{S})$ be a probability kernel satisfying $\lambda_{\sigma\#}\mathbf{r} =_{a.s.} \mathbf{r}$ for all $\lambda_\sigma \in \Lambda$. Then $\mathbf{r}$ admits a unique decomposition*

$$\mathbf{r}(\,.\,, y) = \int_{\mathcal{P}_m} \mathbf{p}(\,.\,, s)\mathbf{q}(ds, y) \;, \tag{3.7}$$

where $\mathbf{p}$ is defined as above and $\mathbf{q} = (|\,.\,|^{\downarrow} \circ \mathrm{part})_{\#}\mathbf{p}$.

If $\mathrm{part}(\,.\,)$ is applied to both sides of (3.7), $\mathrm{part}_{\#}\mathbf{p}$ is a paint-box kernel in the sense of Kingman and $\mathrm{part}_{\#}\mathbf{r}(\,.\,, y)$ the law of an exchangeable partition. See [32] for related results stated in the language of harmonic functions.

3.3. *Random Dense Graphs.* We consider modeling problems where measurements are finite graphs. An observed graph is assumed to be a partial observation of an underlying, larger random graph of possibly infinite size. This random graph is assumed exchangeable, i.e. its distribution is invariant under relabeling of vertices. A suitable parameter space and the requisite Polish topology on this space are obtained by the method of graph limits [38, 39]; see also [5, 17].

EXAMPLE 3.6 (Graphs as projective limits). Let $\mathbf{L}_{[n]}$ be the set of labeled, undirected, simple graphs with vertex set $[n]$. Denote by $\binom{[n]}{2}$ the set of unordered pairs of elements in $[n]$. A graph $x_{[n]} \in \mathbf{L}_{[n]}$ is a pair $x_{[n]} = ([n], E(x_{[n]}))$, where $E(x_{[n]}) \subset \binom{[n]}{2}$ is a set of edges. A random graph $X_{[n]}$ is a graph-valued random variable $X_{[n]} : \Omega \to \mathbf{L}_{[n]}$ with fixed vertex set $[n]$ and random edge set $E(X_{[n]})$. We choose the index set $\Gamma = \{[n] \,|\, n \in \mathbb{N}\}$, ordered by inclusion, and $\mathbf{X}_{[n]} = \mathbf{L}_{[n]}$. Define a projector $\mathbf{L}_{[n+1]} \to \mathbf{L}_{[n]}$ as

$$f_{[\mathrm{n}+1][\mathrm{n}]}(x_{[n+1]}) := ([n], E(x_{[n+1]}) \cap \tbinom{[n]}{2}) \;, \tag{3.8}$$

which deletes vertex $(n+1)$ and all associated edges from the graph $x_{[n+1]}$. A graph $x_{[n]} \in \mathbf{L}_{[n]}$ can be interpreted as a mapping $x_{[n]} : \binom{[n]}{2} \to \{0, 1\}$. Since the spaces $\mathbf{L}_{[n]}$ are product spaces, the projective limit is again a product space, specifically,

$$\mathbf{L}_{[n]} \cong \{0,1\}^{\binom{[n]}{2}} \qquad \text{and} \qquad \mathbf{X}_\Gamma = \varprojlim \langle \mathbf{X}_{[n]} \rangle_\Gamma = \{0,1\}^{\binom{\mathbb{N}}{2}} = \mathbf{L}_{\mathbb{N}} \;. \tag{3.9}$$

We endow $\mathbf{L}_{\mathbb{N}}$ with the product topology and the corresponding Borel sets $\mathcal{B}_\Gamma$, and assume $\mathbf{X} = \mathbf{X}_\Gamma$ and $\mathrm{F} = \mathrm{Id}_{\mathbf{L}_{\mathbb{N}}}$ in (2.2). Graphs obtained as projective limits are called *dense* (resp. *sparse*) if the number of edges scales quadratically (resp. linearly) with the number of nodes as n increases.

Let $\mathcal{W}$ be the set of all measurable, symmetric functions $w : [0,1]^2 \to [0,1]$. Any finite graph $x_{[n]} \in \mathbf{L}$ can be represented by a function $w_{x_{[n]}} \in \mathcal{W}$ as follows: Decompose the unit square into n intervals $A_i^n := ((i-1)/n, i/n]$, and represent the graph by the indicator function

$$w_{x_{[n]}}(t,t') := \sum_{(i,j)\in E(x_{[n]})} \mathbb{I}_{A_i^n \times A_j^n}(t,t') \; . \tag{3.10}$$

Thus, each edge (i,j) corresponds to a patch $A_i^n \times A_j^n$ of the unit square, and the function is non-zero on this patch iff (i,j) is present in the graph. Since $\mathcal{W} \subset L_1[0,1]^2$ and step functions are dense in $L_1[0,1]^2$, the space $(\mathcal{W}, \|\,.\,\|_{L_1})$ can be regarded as the closure of all representations of graphs in $\mathbf{L}_{\mathbb{N}}$ by functions of the form (3.10). Elements of $\mathcal{W}$ are called *graphons* in the literature [17].

Any graphon w parameterizes a random graph X^w with values in $\mathbf{L}_{\mathbb{N}}$. The random set of edges (i,j) of X^w is generated by

$$U_1, U_2, \dots \sim_{\text{iid}} \text{Uniform}[0,1] \quad \text{and} \quad (i,j) \sim \text{Bernoulli}(w(U_i,U_j)) \; . \tag{3.11}$$

The graph X^w is dense unless $w =_{\text{a.s.}} 0$. The parameterization by w is not unique: Different functions w can determine the same random graph X^w [38]. Define an equivalence relation $\equiv$ on $\mathcal{W}$ by $w \equiv w'$ if $X^w \overset{\text{d}}{=} X^{w'}$. We consider the quotient space $\mathbf{W} := \mathcal{W}/\equiv$ and write $[w]_{\equiv}$ for the equivalence class of w in $\mathcal{W}$. The space $\mathbf{W}$ is compact and metrizable [39, Theorem 5.1]. The random graphs X^w can now be re-parameterized on $\mathbf{W}$ as X^{w} with $\mathrm{w} := [w]_{\equiv}$.

Example 3.7 (Exchangeable graph models). By means of (3.10), define graph functionals $\mathrm{w}_{x_{[n]}} := [w_{x_{[n]}}]_{\equiv}$. The functionals are invariant under permutations $\pi \in \mathbb{S}_n$ acting on the vertex labels of a graph and hence play a role similar to that of the empirical measure for sequence data, preserving all information bar label order. A random graph X with values in $\mathbf{X} = \mathbf{L}_{\mathbb{N}}$ is *exchangeable* if its law $P = X_{\#}\mathbb{P}$ satisfies $\pi_{\#}P = P$ for every permutation $\pi \in \mathbb{S}_{\infty}$ acting on the set $\mathbb{N}$ of node labels. Clearly, $\mathrm{w}_{x_{[n]}}$ is a sufficient statistic for the set of exchangeable random graph distributions on $\mathbf{L}_{[n]}$. Denote by $\mathbf{v}_{[n]}$ the corresponding sufficient kernel. In the limit, application of Theorem 2.9 yields a sufficient statistic formulation of a well-known result of Aldous and Hoover [29, Theorem 7.22], which recently has received renewed attention in the context of graph limits [38, 17].

Corollary 3.8. *Define statistics $S_{[n]} : \mathbf{L}_{[n]} \to \mathbf{W}$ as $x_{[n]} \mapsto [w_{x_{[n]}}]_{\equiv}$, and let $\mathbf{v}_{[n]}$ be a sufficient kernel for the exchangeable laws on $\mathbf{L}_{[n]}$ and $S_{[n]}$.*

The family $\langle S_{[n]} \rangle_\Gamma$ *admits a limit in the sense of* (2.22)*, given by*

$$S : \mathbf{L}_\mathbb{N} \to \mathbf{W} \qquad \text{with} \qquad x \mapsto [w_x]_\equiv \; . \tag{3.12}$$

The set $M_{\mathrm{S},\mathbf{v}}$*, where* $\mathbf{v} := \varprojlim \langle \mathbf{v}_{[n]} \rangle_\Gamma$*, is the set of exchangeable random graph distributions on* $\mathbf{L}_\mathbb{N}$*. Its ergodic measures are characterized by* (3.11)*,*

$$\mathcal{E}_{\mathrm{S},\mathbf{v}} = \{ \mathbf{p}(\,.\,,\mathrm{w}) := \mathit{Law}(X^{\mathrm{w}}) \,|\, \mathrm{w} \in \mathbf{W} \} \; . \tag{3.13}$$

Let $\langle \mathbf{r}_{[n]} : \mathbf{Y} \to \mathbf{M}(\mathbf{L}_{[n]}) \rangle_\Gamma$ *be a tight conditional promeasure with limit* $\mathbf{r}$*. If the kernels satisfy* $\pi_\# \mathbf{r}_{[n]} = \mathbf{r}_{[n]}$ *for all permutations* $\pi_{[n]} \in \mathbb{S}_n$*, there exists an a.s.-unique probability kernel* $\mathbf{q} : \mathbf{Y} \to \mathbf{M}(\mathbf{W})$ *such that*

$$\mathbf{r}(\,.\,,y) = \int_\mathbf{W} \mathbf{p}(\,.\,,\mathrm{w}) \mathbf{q}(d\mathrm{w}, y) \; . \tag{3.14}$$

One may now proceed to define nonparametric Bayesian models for graph-valued data, in particular using conditional promeasures $\langle \mathbf{p}_\mathrm{I} \rangle_\Gamma$ and $\langle \mathbf{q}_\mathrm{I} \rangle_\Gamma$ defined in terms of exponential family models and their natural conjugate priors. Theorem 2.7 then guarantees conjugate posteriors with updates of the form (2.15). Several caveats are worth noting: (1) The evaluation of many interesting graph statistics (e.g., is the graph 3-colorable?) is a computationally hard problem; (2) the partition functions of parametric random graph models can be computationally prohibitive; (3) seemingly meaningful graph statistics can behave in an unexpected manner as the graph grows large [11]; and (4) for many statistical problems, graphs should be sparse not dense, but the analogous analytic theory of sparse graphs is still in its early stages.

3.4. *Random Functions.* As a counter point to the discrete flavor of the previous examples, we consider the estimation of a random function on a continuous domain [e.g. 47, 50]. Suppose measurements $x_i \in \mathbb{R}$ are recorded at distinct covariate locations $s_i \in [0,1]$. Each measurement

$$x_i = \theta_i + \varepsilon_i \tag{3.15}$$

consists of a value θ_i corrupted by additive white noise ε_i with variance σ^2. We choose $\mathbf{T} = L_2[0,1]$ and the projective system $\langle \mathbf{T}_\mathrm{I}, g_\mathrm{JI} \rangle_\Gamma = \langle \mathbb{R}^\mathrm{I}, \mathrm{pr}_\mathrm{I} \rangle_\Gamma$. Hence, $\mathbf{T}_\Gamma = \mathbb{R}^\mathbb{N}$. For the embedding G, choose an orthonormal basis $(e_i)_{i \in \mathbb{N}}$ of $L_2[0,1]$, and define $\mathrm{G} : \theta \mapsto (\langle \theta, e_i \rangle)_{i \in \mathbb{N}}$. Each e_i can be interpreted as a sensor, recording measurement $\langle \theta, e_i \rangle$. Specifically, for (3.15), we choose the Dirac system $e_i = \delta_{s_i}$. The mapping G is an isomorphism of the separable Hilbert spaces $L_2[0,1]$ and $\mathrm{G}(\mathbf{T}) = \ell_2$. Equation (3.15) also implies $\mathbf{X}_\mathrm{I} = \mathbb{R}^\mathrm{I}$, where I is a finite set of indices of covariate locations, and $\mathbf{X} = \mathbb{R}^\mathbb{N}$.

EXAMPLE 3.9 (Gaussian process prior). Denote by $\mathbf{TC}([0,1])$ the set of all positive definite Hermitian operators of "trace class" on $L_2[0,1]$, i.e. with finite trace $\mathrm{tr}(\Sigma) < \infty$. As in Example 2.6, fix $\Sigma \in \mathbf{TC}([0,1])$ and define each kernel $\mathbf{q}_{\mathrm{I}} : \mathbb{R}^{\mathrm{I}} \to \mathbf{M}(\mathbb{R}^{\mathrm{I}})$ as the location family of multivariate Gaussian measures with covariance $\Sigma_{\mathrm{I}} = (\mathrm{H}_{\mathrm{I}} \otimes \mathrm{H}_{\mathrm{I}})\Sigma$, where $\mathrm{H}_{\mathrm{I}} = \mathrm{G}_{\mathrm{I}} = \mathrm{pr}_{\mathrm{I}} \circ \mathrm{G}$. To define a white-noise observation model, let Σ^{ε} be the diagonal positive definite operator satisfying $(\mathrm{pr}_{\{\mathrm{i}\}} \otimes \mathrm{pr}_{\{\mathrm{i}\}})\Sigma^{\varepsilon} = \sigma^2$ for all s_i. Hence, $\Sigma^{\varepsilon} \notin \mathbf{TC}([0,1])$ for any $\sigma > 0$. Each kernel $\mathbf{p}_{\mathrm{I}}$ is now defined as the Gaussian location family on $\mathbb{R}^{\mathrm{I}}$ with fixed covariance $(\mathrm{pr}_{\mathrm{I}} \otimes \mathrm{pr}_{\mathrm{I}})(\Sigma^{\varepsilon})$ and random mean Θ_{I}.

An application of Lemma A.1(ii) shows $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ and $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ are conditional promeasures. According to a well-known result on Gaussian processes on Hilbert spaces [e.g. 50, Theorem 3.1], the space ℓ_2 has outer measure 1 under Gaussian process measure on $\mathbb{R}^{\mathbb{N}}$ with mean m and covariance operator Σ if $m \in \ell_2$ and $\Sigma \in \mathbf{TC}([0,1])$. Thus, by Lemma A.2, $\langle \mathbf{q}_{\mathrm{I}} \rangle_{\Gamma}$ is tight on $L_2[0,1]$. By Theorem 2.5, the posterior under a finite number n of observations again concentrates on $L_2[0,1]$. As a projective limit of conjugate exponential families, the model is conjugate by Theorem 2.7 and admits a posterior update of the form (2.15). The marginal posterior indices T_{I} which define T in (2.15) correspond to the exponential family parameterization of the multivariate Gaussian. If they are transformed to the standard parameterization, the resulting posterior index T of the limit model transforms into the more commonly used update equation for Gaussian process posteriors [e.g. 50, Eq. (3.2)].

4. Related work.

We focus on work specifically related to our results and refer to Hjort *et al.* [26] for references on Bayesian nonparametrics.

4.1. *Projective limits in statistics.*

Projective limits are used throughout mathematics to construct infinite-dimensional objects [8, 9]. In probability, they are a standard tool in the construction of stochastic processes [12] and large deviation theory [13]. In Bayesian nonparametrics, projective limits are used implicitly in Gaussian process models, based on the classic work of Kolmogorov and Wiener [47, 1]. Various authors [e.g. 23, 25] have pointed out technical problems arising in the application of projective limit constructions in other settings, for example in the original construction of the Dirichlet process Ferguson [19]. These problems are, in our terminology, caused by neglect of tightness conditions. See [41] for a detailed discussion.

Projective limits of conditional probabilities and their applications in statistics were pioneered by Lauritzen in his work on extremal families and ergodic decompositions [36]. The questions considered by Lauritzen do not require regularity of paths, and hence no notion of tightness; see Remark 2.10

for details on his setting. Although Lauritzen's framework is not Bayesian, both the projective limit representation and Theorem 2.9 should be regarded as a generalization of his ideas to nonparametric priors. By Lemma A.2, a probability kernel $\mathbf{p}$ on $\mathbf{X}$ can be regarded as a pull-back of a kernel $\mathbf{p}_\Gamma$ on $\mathbf{X}_\Gamma$. In the classic terminology of Doob, certain special cases of such pull-backs are called *modifications* [12, Chapter IV.24]. From this perspective, and neglecting minor technicalities, the limit of a tight conditional promeasure can be interpreted as a modification (with paths in $\mathbf{X}$) of the limit of a projective statistical field.

4.2. *Ergodic decompositions.* Research on ergodic decompositions was initiated by de Finetti's theorem [28, Theorem 11.10]. Varadarajan showed that the result generalizes from invariance under permutations to invariance under any locally compact, second countable group [e.g. 29, Theorem A1.4]. Hence, ergodic decompositions are induced by symmetries; see [29] for a probabilist's account. From a statistical point of view, invariance under a transformation group and sufficiency of a statistic both express similar concepts, namely which aspects of data do not distinguish between different measures and hence are irrelevant to the estimation process. A number of authors established that ergodic decompositions can indeed be induced by sufficient statistics rather than a group, see [14, 36, 15] for references. Lauritzen's formulation [36] is arguably the most powerful; our Theorem 2.9 is an adaptation of Lauritzen's ergodic decomposition to Bayesian nonparametric models on embedded spaces. Under a natural regularity condition, parametric exponential family models emerge as a special case [35, 16]. Recent results in descriptive set theory further clarify the relationship between symmetry and sufficiency: Both the orbits of a group or the fibers of a sufficient statistic partition $\mathbf{X}$ into equivalence classes, and it is this partition which induces the integral decomposition [31, Theorem 3.3]. If in particular each fiber of S is a countable set, there exists a countable group with the fibers as its orbits [31, Theorem 1.3].

4.3. *Posterior properties and conjugacy.* Conjugacy has long been used in parametric Bayesian statistics, where it is also known as *closure under sampling.* The distinction between conjugacy and parametric conjugacy is recent [37]—perhaps surprisingly so, considering its importance in Bayesian nonparametrics. The late emergence of this nomenclature is arguably due to the fact that closure under sampling and parametric conjugacy coincide for parametric models, up to pathological cases. Not so in the nonparametric case, where neutral-to-the-right priors [26], for instance, provide a non-trivial example of a model class which is closed under sampling but not parametrically conjugate. However, no systematic study of conjugacy seems to be available in the Bayesian nonparametric literature to date, notwithstanding its widely acknowledged importance [49, 26].

Kim [34] elegantly used conjugacy to derive convergence properties of the posterior. More generally, asymptotic properties of a range of nonparametric posteriors have been clarified in recent years. Examples include Dirichlet process mixtures [24] and Gaussian process models for nonparametric problems [47] and semiparametric problems [10]. See [26, Chapter 2] for an introduction. The weak convergence property (2.6), which is obtained by a Cramér-Wold device for projective limits due to Pollard [44], cannot substitute for such results, because it neglects a null set of the prior.

4.4. *Applications.* The space $\mathfrak{S}$ of virtual permutations in Sec. 3.2 was introduced by Kerov *et al.* [33], motivated by asymptotic representation theory problems. We refer to Arratia *et al.* [4, Chapter 1] for a clear introduction to random permutations and their properties and to Bertoin [7] and Schweinsberg [46] for background on exchangeable random partitions.

The random dense graphs in Sec. 3.3 are based on the construction of Lovász and Szegedy [38, 39]. Austin and Tao [6] develop similar constructions for hypergraphs. In each, versions of a result of Aldous and of Hoover on exchangeable arrays [e.g. 29, Chapter 7] play a key role, although they appear in various shapes and guises. Corollary 3.8 can be regarded as a statistical formulation of this result. Equivalences between [38] and the Aldous-Hoover theorem are established by Diaconis and Janson [17] and Austin [5]. Since all these results can ultimately be traced back to de Finetti's theorem, it is interesting to note how closely the "recipes" formulated in [5, 6] resemble a Bayesian sampling scheme. See [2, 5] for further references.

The construction of random probability measures in Sec. 3.1 follows [41], but is in essence due to Ferguson's construction of the Dirichlet process [19]. The approach is mostly of interest in the context of random measures which, unlike the Dirichlet process, do not admit a stick-breaking construction; see [41] for further examples. Two distinct constructions can be considered for the Gaussian process, depending on whether the limit measure is meant to concentrate on a space of continuous or a space of square-integrable functions. Both constructions are due to the classic work of Kolmogorov, Doob and Wiener; the L_2 construction in Sec. 3.4 follows Zhao [50].

APPENDIX A: PROOFS

We first state some auxiliary results to simplify the work with conditional promeasures; their proofs are deferred until the end of this section. The first lemma helps determine whether a family $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ is a conditional promeasure.

LEMMA A.1. *Let $\langle \mathbf{p}_{\mathrm{I}} \colon \Omega \to \mathbf{M}(\mathbf{X}_{\mathrm{I}}) \rangle_{\Gamma}$ be a family of probability kernels on a projective system $\langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$, and let $X_{\mathrm{I}} : \Omega \to \mathbf{X}_{\mathrm{I}}$ be random variables.*
(i) Let $\langle \mathcal{C}_{\mathrm{I}} \rangle_{\Gamma}$ be a filtration, i.e. $\mathcal{C}_{\mathrm{I}} \subset \mathcal{C}_{\mathrm{J}}$ if $I \preceq J$. If $\mathbf{p}_{\mathrm{I}} =_{a.s.} \mathbb{P}[X_{\mathrm{I}} \in \,.\, | \mathcal{C}_{\mathrm{I}}]$ for all $I \in \Gamma$, then $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ is a conditional promeasure iff

$$f_{\mathrm{JI}} \circ X_{\mathrm{J}} \stackrel{d}{=} X_{\mathrm{I}} \quad \text{and} \quad X_{\mathrm{I}} \perp\!\!\!\perp_{\mathcal{C}_{\mathrm{I}}} \mathcal{C}_{\mathrm{J}} \qquad \text{for all } I \preceq J\,. \tag{A.1}$$

(ii) Let $\mathcal{I}$ be a countable set. For each $i \in \mathcal{I}$, let $\mathbf{X}_{\{\mathrm{i}\}}$ be Polish and $\nu_{\{\mathrm{i}\}}$ a σ-finite measure on $\mathbf{X}_{\{\mathrm{i}\}}$. For each finite $I \subset \mathcal{I}$, let $\mathbf{p}_{\mathrm{I}}$ be a probability kernel on $\mathbf{X}_{\mathrm{I}} := \prod_{\{\mathrm{i}\} \in \mathrm{I}} \mathbf{X}_{\{\mathrm{i}\}}$. If each $\mathbf{p}_{\mathrm{I}}$ has a conditional density $p_{\mathrm{I}}(x_{\mathrm{I}} | \theta_{\mathrm{I}})$ with respect to $\nu_{\mathrm{I}} := \otimes_{\{\mathrm{i}\} \in \mathrm{I}} \nu_{\{\mathrm{i}\}}$, the family $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ is a conditional promeasure iff

$$\int_{\mathbf{X}_{\mathrm{J} \setminus \mathrm{I}}} p_{\mathrm{J}}(x_{\mathrm{J}} | \theta_{\mathrm{J}}) d\nu_{\mathrm{J} \setminus \mathrm{I}}(x_{\mathrm{J} \setminus \mathrm{I}}) =_{a.s.} p_{\mathrm{I}}(x_{\mathrm{I}} | f_{\mathrm{JI}} \theta_{\mathrm{J}}) \tag{A.2}$$

for ν_{I}-almost all $x_{\mathrm{I}} \in \mathbf{X}_{\mathrm{I}}$ and whenever $I \preceq J$.

Tightness can be restated as an outer measure condition. In the next lemma, $\mathbf{p}^*_{\Gamma}(\,.\,, \omega)$ denotes the outer measure defined by the measure $\mathbf{p}_{\Gamma}(\,.\,, \omega)$. Recall that the pull-back $\mathrm{F}^{\#} \mathbf{p}_{\Gamma}(\,.\,, \omega)$ exists if $\mathbf{p}_{\Gamma}(\mathrm{F}\mathbf{X}, \omega) = 1$.

LEMMA A.2. *(i) A conditional promeasure $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ is tight on a Hausdorff space $\mathbf{X}$ if and only if*

$$\mathbf{p}^*_{\Gamma}(\mathrm{F}\mathbf{X}, \omega) = 1 \qquad \mathbb{P}\text{-}a.s. \tag{A.3}$$

In this case, the restriction of $\mathbf{p}$ to the sub-σ-algebra $\mathrm{F}^{-1}\mathcal{B}(\mathbf{X}_{\Gamma})$ of $\mathcal{B}(\mathbf{X})$ coincides with the pull-back $\mathrm{F}^{\#}\mathbf{p}_{\Gamma}$. Conversely, $\mathrm{F}^{\#}\mathbf{p}_{\Gamma}$ determines $\mathbf{p}$ uniquely.
(ii) If $\mathbf{X}$ is Polish, $\mathrm{F}\mathbf{X}$ is measurable in $\mathbf{X}_{\Gamma}$, and $\mathbf{p}^*_{\Gamma}$ reduces to $\mathbf{p}_{\Gamma}$ in (A.3).

It can therefore be convenient to construct $\mathbf{p}$ in two stages, by first constructing $\mathbf{p}_{\Gamma}$ on $\mathbf{X}_{\Gamma}$, followed by a pull-back to $\mathbf{X}$.

Several proofs are considerably simplified by working with parameterized families of random variables, rather than with kernels:

LEMMA A.3. *Let $\mathbf{X}$ and $\mathbf{T}$ be Polish embedded spaces and $\mathbf{p} \colon \mathbf{T} \to \mathbf{M}(\mathbf{X})$ a probability kernel.*
(i) For any probability measure ν_{I} on $\mathbf{X}_{\mathrm{I}}$, there exists a random variable $\xi^{\mathrm{F}}_{\mathrm{I}} \colon \mathbf{X}_{\mathrm{I}} \to \mathbf{X}$ for which $\xi^{\mathrm{F}}_{\mathrm{I}}(x_{\mathrm{I}}) \in \mathrm{F}^{-1}_{\mathrm{I}}(x_{\mathrm{I}})$ holds ν_{I}-a.s.

(ii) There exists a measurable mapping $\tilde{X} : \Omega \times \mathbf{T} \to \mathbf{X}$ which defines a family of random variables

$$X^{\theta}(\omega) := \tilde{X}(\omega, \theta) \quad \textit{with} \quad \textit{Law}(X^{\theta}) = \mathbf{p}(\,.\,, \theta) \quad \textit{for all } \theta \in \mathbf{T}\,. \tag{A.4}$$

This result is valid for any measurable space $\mathbf{T}$.
(iii) Let $X : \Omega \to \mathbf{X}$ and $X_{\mathrm{I}} : \Omega \to \mathbf{X}_{\mathrm{I}}$ be random variables. If $\mathrm{F}_{\mathrm{I}}X \stackrel{d}{=} X_{\mathrm{I}}$, there is a random variable $X'_{\mathrm{I}} : \Omega \to \mathbf{X}_{\mathrm{I}}$ such that

$$\mathrm{F}_{\mathrm{I}}X =_{a.s.} X'_{\mathrm{I}} \qquad \textit{and} \qquad X'_{\mathrm{I}} \stackrel{d}{=} X_{\mathrm{I}}\,. \tag{A.5}$$

The same holds mutatis mutandis for $X_{\mathrm{J}} : \Omega \to \mathbf{X}_{\mathrm{J}}$ if $f_{\mathrm{JI}}X_{\mathrm{J}} \stackrel{d}{=} X_{\mathrm{I}}$.
(iv) Let $\mathbf{p} = \mathrm{F}^{\#} \varprojlim \langle \mathbf{p}_{\mathrm{I}} \circ \mathrm{G}_{\mathrm{I}} \rangle_{\Gamma}$ for kernels $\mathbf{p}_{\mathrm{I}} : \mathbf{T}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$. Then the corresponding mappings $\tilde{X}$ and $\tilde{X}_{\mathrm{I}}$ in (A.4) can be chosen such that

$$\mathrm{F}_{\mathrm{I}} \circ X^{\theta} =_{a.s.} \tilde{X}_{\mathrm{I}}(\,.\,, \mathrm{G}_{\mathrm{I}}\theta) \qquad \textit{for all } I \in \Gamma\,. \tag{A.6}$$

A.1. Proof of Theorem 2.3. Condition (2.4) holds for all $\varepsilon > 0$ if and only if it holds for a dense sequence $\varepsilon_i \searrow 0$ in $\mathbb{R}_+$. For each ε_i, let N^i_{ε} be the null set of exceptions. Similarly, for each pair $I \preceq J$ in the countable index set Γ, let N_{JI} be the null set of exceptions up to which $f_{\mathrm{JI}\#}\mathbf{p}_{\mathrm{J}} =_{\mathrm{a.s.}} \mathbf{p}_{\mathrm{I}}$ in (2.3) holds. Denote the aggregate null set by $N := \bigcup_i N^i_{\varepsilon} \cup \bigcup_{I \preceq J} N_{\mathrm{JI}}$.
(1) Construction of $\mathbf{p}$. Prokhorov's theorem on projective limits of measures [9, IX.4.2, Theorem 1] is applicable point-wise in $\omega \in \Omega' \smallsetminus N$. Hence, there exists for fixed ω a unique Radon measure $\mathbf{p}(\,.\,, \omega)$ on $\mathbf{X}$ satisfying $\mathrm{F}_{\mathrm{I}\#}\mathbf{p}(\,.\,, \omega) = \mathbf{p}_{\mathrm{I}}(\,.\,, \omega)$ for all I, i.e. a measure-valued function $\mathbf{p} : \Omega' \smallsetminus N \to \mathbf{M}(\mathbf{X})$ satisfying (2.5). We must extend $\mathbf{p}$ to Ω'. Prokhorov's theorem additionally guarantees

$$\mathbf{p}(K, \omega) = \inf_{I \in \Gamma} \mathbf{p}_{\mathrm{I}}(\mathrm{F}_{\mathrm{I}}K, \omega) \qquad \omega \in \Omega' \smallsetminus N\,, \tag{A.7}$$

for any fixed compact set $K \subset \mathbf{X}$ (see [9, IX.4.2, Theorem 1] and [9, IX.1.2, Corollary 2 of Proposition 7], and note that $\mathrm{F}_{\mathrm{I}}K$ is compact and hence measurable). To extend the function $\mathbf{p}(K, \,.\,)$ from $\Omega' \smallsetminus N$ to Ω', define $\mathbf{p}(K, \omega) := \inf_{\mathrm{I}} \mathbf{p}_{\mathrm{I}}(\mathrm{F}_{\mathrm{I}}K, \omega)$ for each $\omega \in N$ if K is compact. On a Hausdorff space, a probability measure is completely determined by its values on compact sets [9, IX.1.2, Corollary of Proposition 7], which completes the construction.
(2) Measurability. Since each $\mathbf{p}_{\mathrm{I}} : \Omega' \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ is $\mathcal{C}_{\mathrm{I}}$-measurable, (2.3) implies $f_{\mathrm{JI}\#}\mathbf{p}_{\mathrm{J}}$ is measurable with respect to the $\mathbb{P}$-completion $\overline{\mathcal{C}_{\mathrm{I}}}$ [28, Lemma 1.25]. Hence, again by (2.3), the family $(\overline{\mathcal{C}_{\mathrm{I}}})_{\mathrm{I}}$ is a filtration indexed by Γ. Since further $\overline{\mathcal{C}_{\mathrm{T}}} = \limsup_{\mathrm{I}} \overline{\mathcal{C}_{\mathrm{I}}}$ [21, Proposition 7.2.4] and $\overline{\mathcal{C}_{\Gamma}} = \sigma(\overline{\mathcal{C}_{\mathrm{I}}}; I \in \Gamma)$,

all $\mathbf{p}_{\mathrm{I}}$ are measurable with respect to $\overline{\mathcal{C}_\Gamma} = \overline{\mathcal{C}_{\mathrm{T}}}$. Therefore, the infimum $\omega \mapsto \mathbf{p}^*(\mathrm{F}_{\mathrm{I}}K, \omega)$ in (A.7) is $\overline{\mathcal{C}_\Gamma}$-measurable for each compact K. Since the compact sets determine $\mathbf{p}(\,.\,,\omega)$ unambiguously, $\mathbf{p}$ is a $\overline{\mathcal{C}_\Gamma}$-measurable probability kernel $\Omega' \to \mathbf{M}(\mathbf{X})$ satisfying (2.5), and unique outside the null set N. Hence, it has both a $\mathcal{C}_\Gamma$-measurable and a $\mathcal{C}_{\mathrm{T}}$-measurable version [28, Lemma 1.25].

(3) Radon values. For $\omega \notin N$, the measure $\mathbf{p}(\,.\,,\omega)$ is a Radon measure by Prokhorov's theorem [see 22, 418M]. $\square$

A.2. Proof of Theorem 2.5. *Part (i).* We first observe there are random variables $X : \Omega \to \mathbf{X}$ and $\Theta : \Omega \to \mathbf{T}$ such that

$$\text{(A.8)} \qquad \mathbf{p} =_{\text{a.s.}} \mathbb{P}[X \in \,.\, |\Theta] \quad \text{and} \quad \mathbf{q} =_{\text{a.s.}} \mathbb{P}[\Theta \in \,.\, |Y]\;.$$

To verify (A.8), denote the mappings guaranteed by Lemma A.3(ii) as $\tilde{\Theta}\colon \Omega \times \mathbf{Y} \to \mathbf{M}(\mathbf{T})$ for $\mathbf{q}$ and as $\tilde{X}$ for $\mathbf{p}$, and choose

$$\text{(A.9)} \quad X(\omega) := \tilde{X}(\omega, \tilde{\Theta}(\omega, y)) \quad \text{and} \quad \Theta(\omega) := \tilde{\Theta}(\omega, y) \qquad \text{for } y \in \mathbf{Y}\;.$$

Part (i). Fix any $y \in \mathbf{Y}$. Let X and Θ be random variables corresponding to the kernels $\mathbf{p}$ and $\mathbf{q}(\,.\,,y)$ as in (A.9), and define $\mu^y := \mathrm{Law}(X, \Theta)$. Thus, $\widehat{\mathbf{q}}^y$ is a version of $\mu^y[\Theta \in \,.\, |X]$. Now

$$\text{(A.10)} \quad \mathrm{G}_{\mathrm{I}\#}\mu^y[\Theta \in \,.\, |X] = \mu^y[\Theta \in \mathrm{G}_{\mathrm{I}}^{-1}\,.\, |\mathrm{F}_{\mathrm{I}}X, X] = \mathbb{P}[\Theta_{\mathrm{I}} \in \,.\, |X_{\mathrm{I}}, X]\;,$$

where the last identity holds since $(\mathrm{F}_{\mathrm{I}}X, \mathrm{G}_{\mathrm{I}}\Theta) \stackrel{\mathrm{d}}{=} (X_{\mathrm{I}}, \Theta_{\mathrm{I}})$ by construction. By Lemma A.1(i), $\mathbb{P}[\Theta_{\mathrm{I}} \in \,.\, |X_{\mathrm{I}}, X] = \mathbb{P}[\Theta_{\mathrm{I}} \in \,.\, |X_{\mathrm{I}}]$, and hence

$$\text{(A.11)} \qquad \mathbb{P}[\mathrm{G}_{\mathrm{I}}\Theta \in \,.\, |\mathrm{F}_{\mathrm{I}}X = x_{\mathrm{I}}] =_{\text{a.s.}} \widehat{\mathbf{q}}(\,.\,, x_{\mathrm{I}}, y)\;.$$

Therefore, $\widehat{\mathbf{q}}^y(\mathrm{G}_{\mathrm{I}}^{-1}\,.\,, x) =_{\text{a.s.}} \widehat{\mathbf{q}}_{\mathrm{I}}(\,.\,, \mathrm{F}_{\mathrm{I}}x, y)$, which by Theorem 2.3 implies that $\langle \widehat{\mathbf{q}}_{\mathrm{I}} \rangle_\Gamma$ is tight on $\mathbf{T}$. By a.s.-uniqueness of the limit, $\widehat{\mathbf{q}}^y(\,.\,, x) =_{\text{a.s.}} \widehat{\mathbf{q}}(\,.\,, x, y)$.

Part (ii). Abbreviate $\mu_n := \widehat{\mathbf{q}}^y(\,.\,, X^{(\mathrm{n})})$ and $\mu := \widehat{\mathbf{q}}^y(\,.\,, X)$. We have to show weak convergence $\mu_n \xrightarrow{w} \mu$ holds almost surely under the law of X. A result of Pollard [44, Theorem 1] shows that weak convergence $\mathrm{G}_{\mathrm{I}\#}\mu_n \xrightarrow{w} \mathrm{G}_{\mathrm{I}\#}\mu$ of the marginal sequences on all space $\mathbf{T}_{\mathrm{I}}$, $I \in \Gamma$, implies weak convergence on $\mathbf{T}$. The hypotheses of Pollard's results—mild topological properties and τ-additivity of the limit measure—are satisfied since $\mathbf{T}$ is Polish.

Since $\langle \widehat{\mathbf{q}}_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure, $\mathrm{G}_{\mathrm{I}\#}(\widehat{\mathbf{q}}^y(\,.\,, x)) =_{\text{a.s.}} (\mathrm{G}_{\mathrm{I}\#}\widehat{\mathbf{q}}^y)(\,.\,, \mathrm{F}_{\mathrm{I}}x)$. Therefore, $\mathrm{F}_{\mathrm{I}_n}X^{(\mathrm{n})} =_{\text{a.s.}} x_{\mathrm{I}_n} =_{\text{a.s.}} \mathrm{F}_{\mathrm{I}_n}X$ implies

$$\text{(A.12)} \qquad \forall m \le n : \qquad \mathrm{G}_{\mathrm{I}_m\#}\mu_n = \mathrm{G}_{\mathrm{I}_m\#}\mu \qquad \mathbb{P}\text{-a.s.}$$

Since (I_n) is cofinal, there exists for each I a sufficiently large m such that $I \preceq I_m$, and hence a sufficiently large n such that $\mathrm{G}_{\mathrm{I}\#}\mu_n =_{\text{a.s.}} \mathrm{G}_{\mathrm{I}\#}\mu$. In particular, $\mathrm{G}_{\mathrm{I}\#}\mu_n \xrightarrow{w} \mathrm{G}_{\mathrm{I}\#}\mu$ weakly in $\mathbf{M}(\mathbf{T}_{\mathrm{I}})$ a.s. for all $I \in \Gamma$, and almost sure weak convergence $\mu_n \xrightarrow{w} \mu$ follows by [44, Theorem 1]. $\square$

A.3. Proof of Theorem 2.7. In the following, let X^θ and $X_{\mathrm{I}}^{\theta_{\mathrm{I}}}$ denote the variables (A.4) for $\mathbf{p}$ and $\mathbf{p}_{\mathrm{I}}$. Hence, S is sufficient for $\mathbf{p}$ with sufficient kernel $\mathbf{v}$ iff $\mathbb{P}[X^\theta \in \,.\,|S = s] =_{\text{a.s.}} \mathbf{v}(\,.\,, s)$.

Part (i). Suppose first that $(S_{\mathrm{I}}, \mathbf{v}_{\mathrm{I}})$ are sufficient for $\mathbf{p}_{\mathrm{I}}$ for each $I \in \Gamma$. Define probability kernels $\mathbf{p}'_{\mathrm{I}}$ and $\mathbf{q}'$ as

$$\mathbf{p}'_{\mathrm{I}}(\,.\,, x) := \mathbb{P}[S_{\mathrm{I}} \in \,.\,|X_{\mathrm{I}} = x_{\mathrm{I}}] = \delta_{S_{\mathrm{I}}(x_{\mathrm{I}})}(\,.\,) \quad \text{and} \quad \mathbf{q}'_{\mathrm{I}} := \mathbf{p}_{\mathrm{I}}\ . \tag{A.13}$$

Then (2.9) implies that $\langle \mathbf{p}'_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure and tight on $\mathbf{S}$, with limit $\mathbf{p}'(\,.\,, x) = \delta_{S(x)}(\,.\,)$. Now regard $(\mathbf{p}'_{\mathrm{I}}, \mathbf{q}')$ as a Bayesian model and apply Theorem 2.5, with $(\mathbf{p}'_{\mathrm{I}}, \mathbf{q}'_{\mathrm{I}})$ substituted for $(\mathbf{p}_{\mathrm{I}}, \mathbf{q}_{\mathrm{I}})$: The posterior $\widehat{\mathbf{q}}'_{\mathrm{I}}$ is

$$\widehat{\mathbf{q}}'_{\mathrm{I}}(\,.\,, s_{\mathrm{I}}, \theta_{\mathrm{I}}) =_{\text{a.s.}} \mathbb{P}[X_{\mathrm{I}} \in \,.\,|S_{\mathrm{I}} = s_{\mathrm{I}}, \Theta_{\mathrm{I}} = \theta_{\mathrm{I}}] =_{\text{a.s.}} \mathbf{v}_{\mathrm{I}}(\,.\,, s_{\mathrm{I}})\ . \tag{A.14}$$

Since $\langle \mathbf{v}_{\mathrm{I}} \circ \mathrm{E}_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure by hypothesis, it is tight on $\mathbf{X}$ by Theorem 2.5(i). Hence, $\mathbf{v} := \mathrm{F}^{\#} \varprojlim \langle \mathbf{v}_{\mathrm{I}} \circ \mathrm{E}_{\mathrm{I}} \rangle_\Gamma$ is a conditional probability of X given S. Since $\mathbf{v}(\,.\,, s) =_{\text{a.s.}} \widehat{\mathbf{q}}(\,.\,, s, \theta)$ by a.s.-uniqueness in Theorem 2.3, S is sufficient for $\mathbf{p}$ with sufficient kernel $\mathbf{v}$.

Conversely, suppose $(S, \mathbf{v})$ is sufficient for the limit model $\mathbf{p}$. We must construct a sufficient kernel $\mathbf{v}_{\mathrm{I}}$ for each $\mathbf{p}_{\mathrm{I}}$ and S_{I}. By Lemma A.3(i), there are measurable mappings $\xi^{\mathrm{E}}_{\mathrm{I}} : \mathbf{S}_{\mathrm{I}} \to \mathbf{S}$ and $\xi^{\mathrm{G}}_{\mathrm{I}} : \mathbf{T}_{\mathrm{I}} \to \mathbf{T}$ with $\xi^{\mathrm{E}}_{\mathrm{I}}(x_{\mathrm{I}}) \in \mathrm{F}_{\mathrm{I}}^{-1}\{x_{\mathrm{I}}\}$ and $\xi^{\mathrm{G}}_{\mathrm{I}}(\theta_{\mathrm{I}}) \in \mathrm{G}_{\mathrm{I}}^{-1}\{\theta_{\mathrm{I}}\}$. Define $\mathbf{v}_{\mathrm{I}}(\,.\,, s_{\mathrm{I}}) := \mathbf{v}(\mathrm{F}_{\mathrm{I}}^{-1}\,.\,, \xi^{\mathrm{E}}_{\mathrm{I}}(s_{\mathrm{I}}))$. Then $\mathbf{v}_{\mathrm{I}}$ is a measurable mapping $\mathbf{S}_{\mathrm{I}} \to \mathbf{M}(\mathbf{X}_{\mathrm{I}})$ and satisfies $\mathbf{v}_{\mathrm{I}}(\,.\,, \mathrm{E}_{\mathrm{I}} s) = \mathrm{F}_{\mathrm{I}\#}\mathbf{v}(\,.\,, s)$ as required. Now condition $\mathbf{p}_{\mathrm{I}}$ on the limit S. Since $X_{\mathrm{I}}^{\theta_{\mathrm{I}}}(\omega) =_{\text{a.s.}} \mathrm{F}_{\mathrm{I}} \tilde{X}(\omega, \xi^{\mathrm{G}}_{\mathrm{I}}(\theta_{\mathrm{I}}))$ by Lemma A.3(iv), we have

$$\mathbb{P}[X_{\mathrm{I}}^{\theta_{\mathrm{I}}} \in A_{\mathrm{I}} | S = s] =_{\text{a.s.}} \mathbb{P}[\tilde{X}(\,.\,, \xi^{\mathrm{G}}_{\mathrm{I}}(\theta_{\mathrm{I}})) \in \mathrm{F}_{\mathrm{I}}^{-1} A_{\mathrm{I}} | S = s] =_{\text{a.s.}} \mathbf{v}(\mathrm{F}_{\mathrm{I}}^{-1} A_{\mathrm{I}}, s)\ .$$

Therefore, since $S_{\mathrm{I}} = \mathrm{E}_{\mathrm{I}} \circ S$,

$$\mathbb{P}[\tilde{X}(\,.\,, \xi^{\mathrm{G}}_{\mathrm{I}}(\theta_{\mathrm{I}})) \in \mathrm{F}_{\mathrm{I}}^{-1} A_{\mathrm{I}} | \mathrm{E}_{\mathrm{I}} \circ S = s_{\mathrm{I}}] =_{\text{a.s.}} \mathbf{v}_{\mathrm{I}}(A_{\mathrm{I}}, s_{\mathrm{I}})\ . \tag{A.15}$$

Hence, S_{I} is sufficient for $\mathbf{p}_{\mathrm{I}}$ with sufficient kernel $\mathbf{v}_{\mathrm{I}}$.

Part (ii). Suppose first posterior indices T_{I} are given, hence $\widehat{\mathbf{q}}_{\mathrm{I}} = \mathbf{u}_{\mathrm{I}} \circ T_{\mathrm{I}}$ for each I. Therefore, $\langle \mathbf{u}_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure of kernels $\widehat{\mathbf{Y}}_{\mathrm{I}} \to \mathbf{M}(\mathbf{T}_{\mathrm{I}})$

and tight on $\mathbf{T}$. Hence, $\mathbf{u} := \mathrm{G}^{\#} \varprojlim \langle \mathbf{u}_{\mathrm{I}} \rangle_{\Gamma}$ satisfies, for any cylinder set $\mathrm{G}^{-1} A_{\mathrm{I}}$ in $\mathbf{T}$,

$$\begin{aligned} \widehat{\mathbf{q}}(\mathrm{G}_{\mathrm{I}}^{-1} A_{\mathrm{I}}, x, y) =_{\text{a.s.}} \widehat{\mathbf{q}}_{\mathrm{I}}(A_{\mathrm{I}}, \mathrm{F}_{\mathrm{I}} x, \mathrm{H}_{\mathrm{I}} y) &=_{\text{a.s.}} \mathbf{u}_{\mathrm{I}}(A_{\mathrm{I}}, T_{\mathrm{I}}(\mathrm{F}_{\mathrm{I}} x, \mathrm{H}_{\mathrm{I}} y)) \\ &=_{\text{a.s.}} \mathbf{u}(\mathrm{G}_{\mathrm{I}}^{-1} A_{\mathrm{I}}, x, y)\ . \end{aligned} \tag{A.16}$$

Since the cylinder sets determine $\mathbf{u}$ by Lemma A.2, we have $\widehat{\mathbf{q}} =_{\text{a.s.}} \mathbf{u} \circ T$.

Conversely, let T be a posterior index. Let $\xi_{\mathrm{I}}^{\hat{\mathrm{H}}}$ be the mapping guaranteed by Lemma A.3(i) for $\widehat{\mathrm{H}}_{\mathrm{I}}$. Then the kernels $\mathbf{u}_{\mathrm{I}} : \widehat{\mathbf{Y}}_{\mathrm{I}} \to \mathbf{M}(\mathbf{T}_{\mathrm{I}})$ defined by $\mathbf{u}_{\mathrm{I}}(A_{\mathrm{I}}, \hat{y}_{\mathrm{I}}) := \mathbf{u}(\mathrm{G}_{\mathrm{I}}^{-1} A_{\mathrm{I}}, \xi_{\mathrm{I}}^{\hat{\mathrm{H}}} \hat{y}_{\mathrm{I}})$ satisfy $\widehat{\mathbf{q}}_{\mathrm{I}} =_{\text{a.s.}} \mathbf{u}_{\mathrm{I}} \circ T_{\mathrm{I}}$, hence T_{I} is a posterior index for the posterior $\widehat{\mathbf{q}}_{\mathrm{I}}$. $\square$

A.4. Proof of Theorem 2.9. Lauritzen [36] and Diaconis and Freedman [14] give proofs for the cases $\mathbf{X} = \mathbf{X}_{\Gamma}$ and $\mathbf{X} = \mathbf{X}_0^{\mathbb{N}}$, respectively. Rather than adapting their derivations, which proceed from first principles, to the case of embedded spaces, we obtain a concise proof by reduction to results in functional analysis.

Recall a few definitions: Denote by $B_b(\mathbf{X})$ and $C_b(\mathbf{X})$ the spaces of bounded Borel measurable and bounded continuous functions on $\mathbf{X}$. A probability kernel $\mathbf{v} : \mathbf{S} \to \mathbf{M}(\mathbf{X})$ defines a Markov operator $\mathbf{P} \colon B_b(\mathbf{X}) \to B_b(\mathbf{S})$ by means of $(\mathbf{P}f)(s) = \int f(x)\mathbf{v}(dx, s)$. Its norm adjoint $\mathbf{P}' \colon \mathbf{M}(\mathbf{S}) \to \mathbf{M}(\mathbf{X})$ acts on $\nu \in \mathbf{M}(\mathbf{S})$ by means of $(\mathbf{P}'\nu)(A) = \int_{\mathbf{S}} \mathbf{v}(A, s)\nu(ds)$, for $A \in \mathcal{B}(\mathbf{X})$. If $\mu \in \mathbf{M}(\mathbf{X})$ is a probability measure, a measurable set A is called *μ-invariant* if $\mathbf{v}(A, \,.\,) = \mathbb{I}_A(\,.\,)$ μ-a.s. The measure μ is called *$\mathbf{P}$-invariant* if $\mathbf{P}'\mu = \mu$, and *$\mathbf{P}$-ergodic* if $\mu(A) \in \{0, 1\}$ whenever A is μ-invariant. See [3] for an accessible exposition.

Measures in $M_{\mathrm{S},\mathbf{v}}$ are invariant. A measure μ is in $M_{\mathrm{S},\mathbf{v}}$ iff $\mathbf{v}$ is a version of the conditional probability $\mu[\,.\,|S]$, and hence iff $\mu = \int \mathbf{v}(\,.\,, S(x))\mu(dx)$. If we therefore define a kernel $\mathbf{v}_{\mathrm{X}} : \mathbf{X} \to \mathbf{M}(\mathbf{X})$ as $\mathbf{v}_{\mathrm{X}}(\,.\,, x) := \mathbf{v}(\,.\,, S(x))$, and $\mathbf{P}_{\mathrm{X}}$ as its Markov operator, the measures $\mu \in M_{\mathrm{S},\mathbf{v}}$ are precisely the $\mathbf{P}_{\mathrm{X}}$-invariant measures $\mu = \mathbf{P}'_{\mathrm{X}}\mu$.

Convexity and extreme points. A standard ergodic result for metrizable spaces [3, Theorem 19.25] states that the set of $\mathbf{P}_{\mathrm{X}}$-invariant measures of a Markov operator $\mathbf{P}_{\mathrm{X}}$ is convex with the $\mathbf{P}_{\mathrm{X}}$-ergodic measures as its extreme points. Thus, $M_{\mathrm{S},\mathbf{v}}$ is convex and $\mathrm{ex}(M_{\mathrm{S},\mathbf{v}}) = \mathcal{E}_{\mathrm{S},\mathbf{v}}$. What remains to be shown is $\mathcal{E}_{\mathrm{S},\mathbf{v}} = \Delta_{\mathrm{S},\mathbf{v}}$: Suppose first $\eta \in \Delta_{\mathrm{S},\mathbf{v}}$ with $S_{\#}\eta = \delta_{s_\eta}$. Then $\eta = \mathbf{v}(\,.\,, s_\eta)$, and as a conditional probability given S, $\mathbf{v}$ is 0-1 on $\mathcal{S}$. Conversely, let $\eta \in \mathcal{E}_{\mathrm{S},\mathbf{v}}$ and suppose $\eta \notin \Delta_{\mathrm{S},\mathbf{v}}$. Then $S_{\#}\eta$ assigns positive mass to all open neighborhoods of two distinct points $s_1, s_2 \in \mathbf{S}$, therefore $\mathbf{v}(\,.\,, s_1) = \eta = \mathbf{v}(\,.\,, s_2)$, which contradicts $\eta \in \mathrm{ex}(M_{\mathrm{S},\mathbf{v}})$. Hence, (2.20).

τ is a Borel isomorphism. Let $\eta_1, \eta_2 \in \mathcal{E}_{\mathrm{S}}$ be distinct ergodic measures. The set $\{\eta_1, \eta_2\}$ inherits S as a sufficient statistic from $M_{\mathrm{S},\mathbf{v}}$. Consider random variables $X_1 \sim \eta_1$ and $X_2 \sim \eta_2$. Since $\mathbf{S} \times \mathbf{S}$ has measurable diagonal,

$$(\eta_1 \otimes \eta_2)\{S(X_1) \neq S(X_2)\} > 0 \;, \tag{A.17}$$

so $\tau_{\mathcal{E}}$ is injective. Both $\mathbf{S}$ and $\mathbf{M}(\mathbf{X})$ are Polish and $\mathbf{v}\colon \mathbf{S} \to \mathbf{M}(\mathbf{X})$ is measurable. If $\mathbf{v}$ is invertible on a subset $A \in \mathcal{B}(\mathbf{S})$, then by [30, Corollary 15.2], the restriction $\mathbf{v}|_A$ of $\mathbf{v}$ to A is a Borel isomorphism of A and its image $\mathbf{v}(A) \subset \mathbf{M}(\mathbf{X})$. By construction, $\mathbf{v}(\,.\,, \tau(\eta)) = \eta(\,.\,)$, so $\mathbf{T}$ is precisely the preimage $\mathbf{T} = \mathbf{v}^{-1}(\mathcal{E})$ and measurable in $\mathbf{S}$. The restriction $\mathbf{v}|_{\mathbf{T}}$ is injective since $\mathbf{v}|_{\mathbf{T}} = \tau^{-1}$, and τ is therefore a Borel isomorphism. This also implies $\mathbf{T}$ is a standard Borel space [28, Theorem A1.2].

Existence of decomposition. Since $M_{\mathrm{S},\mathbf{v}}$ is not generally compact, we appeal to a non-compact generalization of Choquet's theorem [48, Theorem 1]: The integral decomposition (2.21) exists if (1) $M_{\mathrm{S},\mathbf{v}}$ is convex and weakly closed and (2) all measures in $M_{\mathrm{S},\mathbf{v}}$ are Radon measures [48, Corollary 1]. Weak closure holds by hypothesis and $M_{\mathrm{S},\mathbf{v}}$ contains only Radon measures since $\mathbf{X}$ is Polish. Hence, the representation (2.21) exists for almost all y.

Uniqueness of decomposition. Suppose $\mathbf{q}_1$ and $\mathbf{q}_2$ are two kernels satisfying (2.21). Let Θ_1^y and Θ_2^y be the respective mappings with $\mathrm{Law}(\Theta_i^y) = \mathbf{q}_i(\,.\,, y)$ as guaranteed by Lemma A.3. Then by (2.21), $\mathbf{p} \circ \Theta_1^y \overset{\mathrm{d}}{=} \mathbf{p} \circ \Theta_2^y$. Since $\mathbf{p}$ is injective with inverse τ, we obtain $\mathbf{q}_1 =_{\text{a.s.}} \mathbf{q}_2$. Let additionally $X_{\mathbf{r}}^y$ and $X_{\mathbf{p}}^\theta$ be variables with laws $\mathbf{r}(\,.\,, y)$ and $\mathbf{p}(\,.\,, \theta)$ as in Lemma A.3. Then by definition of $\mathbf{T}$, we have $S \circ X_{\mathbf{p}}^\theta(\omega) = \theta$ for all $\omega \in \Omega$, and hence

$$X_{\mathbf{r}}^y \overset{\mathrm{d}}{=} X_{\mathbf{p}}^{\Theta^y} \quad \Rightarrow \quad S \circ X_{\mathbf{r}}^y \overset{\mathrm{d}}{=} \Theta^y \quad \Rightarrow \quad \mathbf{q} = S_\# \mathbf{r} \;. \qquad \square \tag{A.18}$$

A.5. Proof of Proposition 2.12. We first note that $f \in C_b(\mathbf{X})$ implies $\mathbf{P}_{\mathrm{X}} f \in C_b(\mathbf{X})$: Since S and all kernels $\mathbf{p}_{\mathrm{I}}$ are continuous, the projective limit $\mathbf{v}$ and hence $\mathbf{v}_{\mathrm{X}}$ are also continuous. Consequently, $\mathbf{P}_{\mathrm{X}}$ preserves bounded continuity [3, Theorem 19.14]. We write $\langle f, \mu \rangle := \int f d\mu$. By definition, $\langle \mathbf{P} f, \mu \rangle = \langle f, \mathbf{P}' \mu \rangle$. Suppose a sequence $\mu_n \in M_{\mathrm{S},\mathbf{v}}$ converges weakly to $\mu \in \mathbf{M}(\mathbf{X})$, or equivalently, $\langle f, \mu_n \rangle \xrightarrow{n \to \infty} \langle f, \mu \rangle$ for all $f \in C_b(\mathbf{X})$. Then

$$\langle f, \mathbf{P}'_{\mathrm{X}} \mu \rangle = \langle \mathbf{P}_{\mathrm{X}} f, \mu \rangle = \lim_n \langle \mathbf{P}_{\mathrm{X}} f, \mu_n \rangle = \lim_n \langle f, \mathbf{P}'_{\mathrm{X}} \mu_n \rangle = \lim_n \langle f, \mu_n \rangle = \langle f, \mu \rangle \;.$$

Since $\mathbf{X}$ is metrizable, $\langle f, \mathbf{P}'_{\mathrm{X}} \mu \rangle = \langle f, \mu \rangle$ for all $f \in C_b(\mathbf{X})$ implies $\mathbf{P}'_{\mathrm{X}} \mu = \mu$ [3, Theorem 15.1]. Hence, $M_{\mathrm{S},\mathbf{v}}$ is closed under weak limits. $\square$

A.6. Proof of Proposition 3.3. As a projective limit of finite and hence Polish groups, $\mathfrak{S}$ is Polish. Since $\mathbf{X} = \mathbb{S}_\infty$ is embedded by the inclusion map, we need to know that $\mathbb{S}_\infty$ is Polish in a topology which makes the inclusion mapping continuous. The relative topology ensures continuity since it is generated by the mapping. Since $\psi : \mathfrak{S} \to \prod_{n\in\mathbb{N}}$ is a homeomorphism, the relative topology is equivalently given by the relative topology of $\psi(\mathbb{S}_\infty)$ in $\mathbb{N}^{\mathbb{N}}$, which is just the usual Polish topology on $\mathbb{S}_\infty$ [30, Sec. 9B]. Therefore, Theorem 2.7 is applicable, and the form of the posterior follows immediately if we can show that $\langle \mathbf{p}_n \rangle_\Gamma$ and $\langle \mathbf{q}_n \rangle_\Gamma$ are tight conditional promeasures.
Both families are conditional promeasures. We use Lemma A.1(ii). Let $I := [n]$ and $J := [n+1]$. For $\pi_{\mathrm{I}} = \sigma_{k_1}(1)\cdots\sigma_{k_n}(n)$, the fiber $f_{\mathrm{JI}}^{-1}\pi_{\mathrm{I}}$ consists of all permutations $\sigma_{k_1}(1)\cdots\sigma_{k_n}(n)\sigma_m(n+1)$ with $m \in J$. Fix $\theta_{\mathrm{J}} \in \mathbf{T}_{\mathrm{J}}$ and set $\theta_{\mathrm{I}} := \mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}}$. Then

$$\mathbf{p}_{\mathrm{J}}(f_{\mathrm{JI}}^{-1}\pi_{\mathrm{I}}, \theta_{\mathrm{J}}) = \mathbf{p}_{\mathrm{I}}(\pi_{\mathrm{I}}, \theta_{\mathrm{I}}) \frac{1 + \sum_{m=1}^{n} e^{-\theta^{(n+1)}}}{1 + n e^{-\theta^{(n+1)}}} = \mathbf{p}_{\mathrm{I}}(\pi_{\mathrm{I}}, \theta_{\mathrm{I}})\ . \tag{A.19}$$

Under ψ_{I}, the previous equation becomes $(\mathrm{pr}_{\mathrm{JI}} \circ \psi_{\mathrm{J}})_{\#}\mathbf{p}_{\mathrm{J}}(\,.\,, \theta_{\mathrm{J}}) = \psi_{\mathrm{I}\#}\mathbf{p}_{\mathrm{I}}(\,.\,, \theta_{\mathrm{I}})$, which establishes (A.2). Since each ψ_{I} is a kernel on the product space $\prod_{m\in I}[m]$, Lemma A.1(ii) is applicable and shows $\langle \psi_{\mathrm{I}\#}\mathbf{p}_{\mathrm{I}} \rangle_\Gamma$ is a conditional promeasure. By (3.3), so is $\langle \mathbf{p}_{\mathrm{I}} \rangle_\Gamma$ under f_{JI}. For the priors $\mathbf{q}_{\mathrm{I}}$, which live on the product spaces $\mathbb{R}^{n-1}$, application of Lemma A.1(ii) straightforward.
Tightness of $\langle \mathbf{p}_{\mathrm{I}} \rangle_\Gamma$. A virtual permutation π is in $\mathbb{S}_\infty$ iff it permutes only a finite number of elements, and hence iff $\sum_j W^{(j)}(\pi) < \infty$. If the sum diverges, π contains an infinite cycle. The random variables $W^{(j)}(\pi)$ are independent under the model [20]. By the Borel-Cantelli lemma, the sum converges iff the sum of probabilities $\Pr\{W^{(j)}(\pi) = 1\}$ converges, and hence iff $\theta \in G^{-1}\ell_1(0,1)$.
Tightness of $\langle \mathbf{q}_{\mathrm{I}} \rangle_\Gamma$. It suffices to show $\theta^{(\mathrm{j})} \xrightarrow{j\to\infty} +\infty$ holds almost surely; then $G_j(\theta^{(\mathrm{n})}) \to 0$ at exponential rate by (3.5), and $G(\theta) \in \ell_1$. Since each $\mathbf{q}_{\mathrm{I}}$ is a natural exponential family model, $\gamma^{(\mathrm{j})} = \mathbb{E}[\Theta^{(\mathrm{j})}]$. The assumption $\gamma \in G^{-1}\ell_1$ hence implies $\mathbb{E}[\Theta_\Gamma^{(\mathrm{j})}] > \epsilon$ for any $\epsilon > 0$ and a cofinite number of indices j. The random variables $\Theta^{(\mathrm{j})}$ are independent given (λ, γ). By Kolmogorov's zero-one law, $\mathbb{P}\{(\Theta^{(\mathrm{j})})_j \text{ diverges}\} \in \{0,1\}$. Since the expected sequence diverges, the probability of divergence is non-zero, and $G(\theta) \in \ell_1$ almost surely. ☐

A.7. Proof of Corollary 3.5. To derive the result from Theorem 2.9, let $M \subset \mathbf{M}(\mathfrak{S})$ be the set of all distributions on $\mathfrak{S}$ which are invariant under conjugation. The set M can trivially be regarded as a probability kernel $\mathbf{k} : M \to \mathbf{M}(\mathfrak{S})$, by choosing $\mathbf{k}$ as the canonical injection. Denote by f_n

the canonical mapping $\mathfrak{S} \to \mathbb{S}_n$ given by the projective system, and define $\mathbf{k}_n := f_{n\#}\mathbf{k}$. We first have to show that $\langle \mathbf{k}_n : f_n M \to \mathbf{M}(\mathbb{S}_n)\rangle_\Gamma$ is a conditional promeasure: We use the cycle structure as a sufficient statistic. Let $C_1, \ldots, C_n$ be the (possibly empty) cycles of $\pi_n \in \mathbb{S}_n$, arranged by decreasing size. We define $S_n(\pi_n) := \frac{1}{n}(|C_1|, \ldots, |C_n|)$ on $\mathbb{S}_n$, and $S\colon \pi \mapsto |\operatorname{part}(\pi)|^{\downarrow}$ on $\mathfrak{S}$. Thus, S_n is sufficient for $f_n M$ and depends on π_n only through the induced partition. If $\operatorname{Law}(\pi_n) = \mathbf{k}_n(\,.\,, s_n)$, we therefore have $\pi_n \perp\!\!\!\perp_{S_n} S_{n+1}$, and $\langle \mathbf{k}_n \rangle_\Gamma$ is a conditional promeasure by Lemma A.1(i).

Since the random partitions induced by elements of M are exchangeable, their asymptotic frequencies $\lim_n \frac{1}{n}|B_j \cap [n]|$ exist by Kingman's theorem [7, Theorem 2.1]. In other words, $\langle S_n \rangle_\Gamma$ admits a limit $\lim_{n\to\infty} S_n(f_n\pi) \to S(\pi)$ in the sense of (2.22), and Theorem 2.9 is applicable. The distributions $\mathbf{v}(\,.\,, s)$ are clearly invariant under conditioning on S, and $S(\pi) = s$ almost surely if π is distributed according to $\mathbf{v}(\,.\,, s)$. Since S is sufficient for M, the set M can contain only one distribution with this property for each value of s; since there is a distinct distribution $\mathbf{p}(\,.\,, s)$ in M for each s, the measures $\mathbf{p}(\,.\,, s)$ are the only elements of M invariant under conditioning on S. $\square$

A.8. Proof of Corollary 3.8. The space $\mathbf{S} = \mathbf{W}$ is Polish as required: It is a compact metric space [39, Theorem 5.1], and therefore complete [3, Theorem 3.28] and separable [3, Lemma 3.26].

The set $M_{\mathrm{S},\mathbf{v}}$ consists of the exchangeable laws. The family $\langle S_{[n]} \rangle_\Gamma$ admits a limit since, by definition,

$$\lim_{n\to\infty} S_{[n]}(\mathrm{w}_{x_{[n]}}) = \lim_{n\to\infty} \mathrm{w}_{x_{[n]}} = \mathrm{w}_x = S(x)\ . \tag{A.20}$$

Representations in $\mathbf{W}$ do not distinguish homomorphic graphs, and therefore $S \circ \pi = S$. On the other hand, representations are by definition unique up to homomorphy, hence $S(x) \neq S(x')$ unless x and x' are homomorphic. Hence, $M_{\mathrm{S},\mathbf{v}}$ consists of the exchangeable laws in $\mathbf{M}(\mathbf{L}_{\mathbb{N}})$.

The set $M_{\mathrm{S},\mathbf{v}}$ is weakly closed. To apply Theorem 2.9, we must show $M_{\mathrm{S},\mathbf{v}}$ is closed. To this end, let $P^{(\mathrm{n})} \in M_{\mathrm{S},\mathbf{v}}$ be a sequence of measures which converges weakly to some $P \in \mathbf{M}(\mathbf{L}_{\mathbb{N}})$. We have to show that $\pi_\# P = P$. Since the weak topology is generated by the evaluation functionals F_A for $A \in \mathcal{B}(\mathbf{L}_{\mathbb{N}})$, weak convergence $P^{(\mathrm{n})} \Rightarrow P$ is equivalent to convergence $\mathrm{F}_{A\#}P^{(\mathrm{n})} \to \mathrm{F}_{A\#}P$ in $[0,1]$ for all $A \in \mathcal{B}(\mathbf{L}_{\mathbb{N}})$. Therefore,

$$\lim \mathrm{F}_{A\#}(\pi_\# P^{(\mathrm{n})}) = \lim \mathrm{F}_{\pi A\#} P^{(\mathrm{n})} = \mathrm{F}_{\pi A\#} P = \mathrm{F}_{A\#}(\pi_\# P) \tag{A.21}$$

and hence $P = \lim P^{(\mathrm{n})} = \lim(\pi_\# P^{(\mathrm{n})}) = \pi_\# P$. The limit P is thus again exchangeable, and $M_{\mathrm{S},\mathbf{v}}$ is closed.

Ergodic measures and decomposition. By (2.20), a random graph X with distribution $P \in M_{\mathrm{S},\mathbf{v}}$ is ergodic iff $S(X) =_{\text{a.s.}} \mathrm{w}$ for some $\mathrm{w} \in \mathbf{W}$. The random graphs X^{w} satisfy $S(X^{\mathrm{w}}) =_{\text{a.s.}} \mathrm{w}$ by construction. Conversely, suppose X is exchangeable and $S(X) =_{\text{a.s.}} \mathrm{w}$. Then the laws of X and X^{w} are identical up to homomorphy of samples. Since both are exchangeable, $X \stackrel{\mathrm{d}}{=} X^{\mathrm{w}}$, which shows all ergodic measures are of the form $\mathrm{Law}(X^{\mathrm{w}})$. The integral decomposition (3.14) now follows from (2.21). ☐

A.9. Proof of Lemma A.1. *Part (i).* Suppose first (A.1) holds. By the properties of conditional independence [28, Proposition 6.6],

$$X_{\mathrm{I}} \perp\!\!\!\perp_{\mathcal{C}_{\mathrm{I}}} \mathcal{C}_{\mathrm{J}} \quad \Leftrightarrow \quad \mathbb{P}[A|\mathcal{C}_{\mathrm{I}}, \mathcal{C}_{\mathrm{J}}] =_{\text{a.s.}} \mathbb{P}[A|\mathcal{C}_{\mathrm{I}}] \quad \text{for all } A \in \sigma(X_{\mathrm{I}}) \ . \tag{A.22}$$

By Lemma A.3(iii), $f_{\mathrm{JI}} X_{\mathrm{J}} \stackrel{\mathrm{d}}{=} X_{\mathrm{I}}$ implies we can replace X_{J} by another variable $X'_{\mathrm{J}} \stackrel{\mathrm{d}}{=} X_{\mathrm{J}}$ satisfying $f_{\mathrm{JI}} X'_{\mathrm{J}} =_{\text{a.s.}} X_{\mathrm{I}}$, and the assertion follows:

$$\mathbf{p}_{\mathrm{J}}(f_{\mathrm{JI}}^{-1} A_{\mathrm{I}}, \omega) \; = \; \mathbb{P}[X_{\mathrm{I}}^{-1} A_{\mathrm{I}} | \mathcal{C}_{\mathrm{J}}] \stackrel{(\text{A.22})}{=} \mathbf{p}_{\mathrm{I}}(A_{\mathrm{I}}, \omega) \qquad \mathbb{P}\text{-a.s.} \tag{A.23}$$

Conversely, assume $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ satisfies (2.3), hence $\mathbf{p}_{\mathrm{J}}(f_{\mathrm{JI}}^{-1} A_{\mathrm{I}}, \omega) =_{\text{a.s.}} \mathbf{p}_{\mathrm{I}}(A_{\mathrm{I}}, \omega)$ for all $A_{\mathrm{I}} \in \mathcal{B}(\mathbf{X}_{\mathrm{I}})$. Since the kernels $\mathbf{p}_{\mathrm{I}}$ are conditional distributions of the variables X_{I}, integrating out ω under $\mathbb{P}$ yields $\mathbb{P}\{f_{\mathrm{JI}} X_{\mathrm{I}} \in A_{\mathrm{I}}\} = \mathbb{P}\{X_{\mathrm{I}} \in A_{\mathrm{I}}\}$ for all A_{I}, and therefore $f_{\mathrm{JI}} X_{\mathrm{J}} \stackrel{\mathrm{d}}{=} X_{\mathrm{I}}$.
Part (ii). By substitution into Corollary (2.4), the family $\langle \mathbf{p}_{\mathrm{I}} \rangle_{\Gamma}$ is a conditional promeasure iff the densities satisfy

$$\int_{\mathrm{pr}_{\mathrm{JI}}^{-1} A_{\mathrm{I}}} p_{\mathrm{J}}(x_{\mathrm{J}}|\theta_{\mathrm{J}}) d\nu_{\mathrm{J}}(x_{\mathrm{J}}) = \int_{A_{\mathrm{I}}} p_{\mathrm{I}}(x_{\mathrm{I}}|\mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}}) d\nu_{\mathrm{J}}(x_{\mathrm{J}}) \qquad \text{for } A_{\mathrm{I}} \in \mathcal{B}_{\mathrm{I}} \ , \tag{A.24}$$

and (A.2) follows by Fubini's theorem. ☐

A.10. Proof of Lemma A.2. Only (i) requires proof. Part (ii) then follows immediately: As a countable projective limit of Polish spaces, $\mathbf{X}_{\Gamma}$ is Polish. F is hence an injective, measurable mapping between Polish spaces, which implies its image $\mathrm{F}\mathbf{X}$ is a Borel set in $\mathbf{X}_{\Gamma}$ [42, Theorem I.3.9].

Equation (A.3) implies tightness. Suppose $\mathbf{p}_{\Gamma}^{*}(\mathrm{F}\mathbf{X}, \omega) = 1$ holds for all ω up to a null set N. For $\omega \in \Omega \smallsetminus N$, the measure $\mathbf{p}_{\Gamma}(\,.\,, \omega)$ has a uniquely defined pull-back measure on $\mathrm{F}^{-1}\mathcal{B}_{\Gamma}$ [22, 234F]. In general, $\mathrm{F}^{-1}\mathcal{B}_{\Gamma}$ is only a sub-σ-algebra of $\mathcal{B}(\mathbf{X})$. However, the proof of Prokhorov's extension theorem shows that, if the mappings F_{I} satisfy the conditions (2.1), then any measure on $\mathrm{F}^{-1}\mathcal{B}_{\Gamma}$ has a unique extension to a measure on $\mathcal{B}(\mathbf{X})$ (see e.g. parts

(a) and (d) of the proof of Theorem 418M in [22]). Hence, $\mathrm{F}^{\#}\mathbf{p}_{\Gamma}(\,.\,,\omega)$ defines a unique probability measure $\mathbf{p}(\,.\,,\omega)$ on $\mathcal{B}(\mathbf{X})$. Clearly, $\mathrm{F}_{\mathrm{I}\#}\mathbf{p}(\,.\,,\omega) = f_{\mathrm{I}\#}\mathbf{p}_{\Gamma}(\,.\,,\omega) = \mathbf{p}_{\mathrm{I}}(\,.\,,\omega)$. By Prokhorov's theorem [9, IX.4.2, Theorem 1], such a measure only exists if the family of measures $\langle\mathbf{p}_{\mathrm{I}}(\,.\,,\omega)\rangle_{\Gamma}$ satisfies (2.4). Since this holds whenever $\omega \notin N$, the conditional promeasure $\langle\mathbf{p}_{\mathrm{I}}\rangle_{\Gamma}$ is tight.

Tightness implies (A.3). Suppose $\langle\mathbf{p}_{\mathrm{I}}\rangle_{\Gamma}$ is tight. The outer measure condition $\mathbf{p}^*_{\Gamma}(\mathrm{F}\mathbf{X},\omega) = 1$ is equivalent to

$$\forall A \in \mathcal{B}_{\Gamma}: \quad \mathbf{p}^*_{\Gamma}(A,\omega) = 0 \qquad \text{whenever } A \subset \mathbf{X}_{\Gamma} \smallsetminus \mathrm{F}\mathbf{X}\,, \tag{A.25}$$

see [22, 132E]. Fix $\omega \in \Omega$. Let $K \subset \mathrm{F}\mathbf{X}$ be any compact set and write $C_{\mathrm{I}} := \mathbf{X}_{\mathrm{I}} \smallsetminus \mathrm{F}_{\mathrm{I}}K$. Then $A \subset \mathbf{X}_{\Gamma} \smallsetminus \mathrm{F}\mathbf{X}$ implies $A \subset \cup_{\mathrm{I}} f_{\mathrm{I}}^{-1}C_{\mathrm{I}}$. We observe two properties of the sets $f_{\mathrm{I}}^{-1}C_{\mathrm{I}}$: (1) They are open, since F_{I} is continuous, which makes $\mathrm{F}_{\mathrm{I}}K$ compact. (2) They satisfy $f_{\mathrm{I}}^{-1}C_{\mathrm{I}} \subset f_{\mathrm{J}}^{-1}C_{\mathrm{J}}$ whenever $I \preceq J$, because $I \preceq J$ implies $f_{\mathrm{J}}^{-1}\mathrm{F}_{\mathrm{J}}K \subset f_{\mathrm{I}}^{-1}\mathrm{F}_{\mathrm{I}}K$. On a family of sets with properties (1) and (2), any outer measure P^* satisfies

$$P^*(\cup_{\mathrm{I}} f_{\mathrm{I}}^{-1}C_{\mathrm{I}}) = \sup_{\mathrm{I}} P^*(f_{\mathrm{I}}^{-1}C_{\mathrm{I}}) \tag{A.26}$$

by [9, IX.1.6, Cor. of Prop. 5]. For any $\varepsilon > 0$, there is by hypothesis a compact set $K \subset \mathbf{X}$ satisfying (2.4). By (A.26), for almost all ω,

$$\begin{aligned} \mathbf{p}^*_{\Gamma}(A,\omega) \leq \mathbf{p}^*_{\Gamma}(\cup_{\mathrm{I}} f_{\mathrm{I}}^{-1}(\mathbf{X}_{\mathrm{I}} \smallsetminus \mathrm{F}_{\mathrm{I}}K),\omega) &= \sup_{\mathrm{I}} \mathbf{p}^*_{\Gamma}(f_{\mathrm{I}}^{-1}(\mathbf{X}_{\mathrm{I}} \smallsetminus \mathrm{F}_{\mathrm{I}}K),\omega) \\ &= \sup_{\mathrm{I}} \mathbf{p}^*_{\mathrm{I}}(\mathbf{X}_{\mathrm{I}} \smallsetminus \mathrm{F}_{\mathrm{I}}K,\omega) \leq \varepsilon\,. \end{aligned} \tag{A.27}$$

Hence, $\mathbf{p}_{\Gamma}(A,\omega) =_{\text{a.s.}} 0$ whenever $A \cap \mathrm{F}\mathbf{X} = \emptyset$, and (A.3) holds. ☐

A.11. Proof of Lemma A.3. *(i)* If the graph $\mathrm{gr}(\mathrm{F}_{\mathrm{I}}) \subset \mathbf{X}_{\mathrm{I}} \times \mathbf{X}$ of F_{I} is measurable, then by [28, Theorem A1.4], $\xi^{\mathrm{F}}_{\mathrm{I}}$ exists and is well-defined on the image $\mathrm{pr}_{\mathrm{I}}\mathrm{gr}(\mathrm{F}_{\mathrm{I}})$ under projection onto $\mathbf{X}_{\mathrm{I}}$. As a measurable mapping between Polish spaces, F_{I} has a $\mathcal{B}_{\mathrm{I}} \otimes \mathcal{B}(\mathbf{X})$-measurable graph [3, Theorem 12.28]. Since F_{I} is surjective, $\mathrm{pr}_{\mathrm{I}}\mathrm{gr}(\mathrm{F}_{\mathrm{I}}) = \mathbf{X}_{\mathrm{I}}$, and hence ξ^{F} is well-defined on the entire space $\mathbf{X}_{\mathrm{I}}$.

(ii) is a result given by Kallenberg [28, Lemma 3.22].

(iii) is an application of [28, Corollary 6.11].

(iv) Since $\mathbf{p} = \mathrm{F}^{\#}\varprojlim\langle\mathbf{p}_{\mathrm{I}}\rangle_{\Gamma}$, the relations (2.5) imply $\mathrm{F}_{\mathrm{I}} \circ X^{\theta} \overset{\mathrm{d}}{=} \tilde{X}_{\mathrm{I}}(\,.\,,\mathrm{G}_{\mathrm{I}}\theta)$, and (A.6) follows from (ii). ☐

APPENDIX B: PROJECTIVE LIMITS

Projective limits (or *inverse limits*) are widely used in pure mathematics and probability. Standard references are [8, 9, 12].

Index sets. Dimensions are indexed by a countably infinite, partially ordered, directed index set $(\Gamma, \preceq)$. Here, *directed* means that, for any two $I, J \in \Gamma$ there is a $K \in \Gamma$ for which $I \preceq K$ and $J \preceq K$. A sequence $I_1, I_2, \dots \in \Gamma$ is *cofinal* if, for any $J \in \Gamma$, there is an $n \in \mathbb{N}$ with $J \preceq I_n$. Since Γ is countable, it contains a cofinal sequence.

Projective limits of spaces. Let $\langle \mathbf{X}_{\mathrm{I}} \rangle_{\Gamma}$ be a system of topological spaces indexed by Γ. Let $f_{\mathrm{JI}} : \mathbf{X}_{\mathrm{J}} \to \mathbf{X}_{\mathrm{I}}$, for $I \preceq J$, be continuous, surjective mappings satisfying $f_{\mathrm{II}} = \mathrm{Id}_{\mathbf{T}_{\mathrm{I}}}$ and $f_{\mathrm{KI}} = f_{\mathrm{JI}} \circ f_{\mathrm{KJ}}$ whenever $I \preceq J \preceq K$. We call the functions f_{JI} *generalized projections* and the collection $\langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$ a *projective system*. Directed sets generalize sequences $(x_n)_n$ to nets $\langle x_{\mathrm{I}} \rangle_{\Gamma}$, i.e. to mappings $I \mapsto x_{\mathrm{I}} \in \mathbf{X}_{\mathrm{I}}$ from indices to points. The set of all nets is the product set $\prod_{I \in \Gamma} \mathbf{X}_{\mathrm{I}}$. The subset $\mathbf{X}_{\Gamma} := \varprojlim \langle \mathbf{X}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\Gamma}$ of all nets satisfying $f_{\mathrm{JI}} x_{\mathrm{J}} = x_{\mathrm{I}}$ whenever $I \preceq J$ is called the *projective limit set*. If $x_{\Gamma} := \langle x_{\mathrm{I}} \rangle_{\Gamma}$ is an element of $\mathbf{X}_{\Gamma}$, the assignment $f_{\mathrm{I}} : x_{\Gamma} \mapsto x_{\mathrm{I}}$ is a well-defined mapping $\mathbf{X}_{\Gamma} \to \mathbf{X}_{\mathrm{I}}$, called a *canonical mapping*. If each space $\mathbf{X}_{\mathrm{I}}$ carries a topology $\mathrm{Top}_{\mathrm{I}}$ and a Borel σ-algebra $\mathcal{B}_{\mathrm{I}}$, the canonical mappings induce a *projective limit topology* Top_{Γ} and a *projective limit σ-algebra* $\mathcal{B}_{\Gamma}$ on $\mathbf{X}_{\Gamma}$. These are the smallest topology and σ-algebra which make all canonical mappings continuous and measurable, respectively. They satisfy $\mathcal{B}_{\Gamma} = \sigma(\mathrm{Top}_{\Gamma})$. The *projective limit space* is the set $\mathbf{X}_{\Gamma}$ endowed with Top_{Γ} and $\mathcal{B}_{\Gamma}$ [8, I.4.4]. The projective limit of a sequence of Polish spaces—i.e. Γ is countable and totally ordered—is Polish [30, §17D]. Since any countable directed set contains a cofinal sequence, a countable projective limit of Polish spaces is also Polish.

Projective limits of mappings and measures. Let $\langle \mathbf{X}'_{\mathrm{I}}, f'_{\mathrm{JI}} \rangle_{\Gamma}$ be a second projective system also indexed by Γ. If $w_{\mathrm{I}} \colon \mathbf{X}_{\mathrm{I}} \to \mathbf{X}'_{\mathrm{I}}$ is a family of measurable mappings with $w_{\mathrm{I}} \circ f_{\mathrm{JI}} = f'_{\mathrm{JI}} \circ w_{\mathrm{J}}$ whenever $I \preceq J$, there is a unique and measurable mapping $w_{\Gamma} \colon \mathbf{X}_{\Gamma} \to \mathbf{X}'_{\Gamma}$ satisfying $w_{\mathrm{I}} \circ f_{\mathrm{I}} = f'_{\mathrm{I}} \circ w_{\Gamma}$. Such a family is called a *projective family* of mappings with *projective limit* w_{Γ}. If each w_{I} is injective, or bijective, or continuous, then so is the projective limit w_{Γ} [8].Similarly, any family $\langle P_{\mathrm{I}} \rangle_{\Gamma}$ of probability measures satisfying the equations $f_{\mathrm{JI}}(P_{\mathrm{J}}) = P_{\mathrm{I}}$ is called a projective family, and has a unique projective limit measure P_{Γ} on $(\mathbf{X}_{\Gamma}, \mathcal{B}(\mathbf{X}_{\Gamma}))$ satisfying $f_{\mathrm{I}}(P_{\Gamma}) = P_{\mathrm{I}}$ for all I [9, IX.4.3, Theorem 2]. To emphasize the fact that the measures P_{I} completely defines P_{Γ}, a projective family $\langle P_{\mathrm{I}} \rangle_{\Gamma}$ is also called the *weak distribution* of a random variable $X_{\Gamma} \sim P_{\Gamma}$, or a *promeasure* [9].

REFERENCES

[1] Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer.

[2] Aldous, D. J. (2010). Exchangeability and continuum limits of discrete random structures. In *Proceedings of the International Congress of Mathematicians*.

[3] Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis*. Springer, 3rd edition.
[4] Arratia, R., Barbour, A. D., and Tavaré, S. (2003). *Logarithmic combinatorial structures: A probabilistic approach*. European Mathematical Society.
[5] Austin, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, **5**, 80–145.
[6] Austin, T. and Tao, T. (2010). Testability and repair of hereditary hypergraph properties. *Random Structures & Algorithms*, **36**, 373–463.
[7] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge University Press.
[8] Bourbaki, N. (1966). *Elements of Mathematics: General Topology*. Hermann (Paris) and Addison-Wesley.
[9] Bourbaki, N. (2004). *Elements of Mathematics: Integration*. Springer.
[10] Castillo, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields*, **152**, 53–99.
[11] Chatterjee, S. and Diaconis, P. (2011). Estimating and understanding exponential random graph models. Preprint. arXiv 1102.2650.
[12] Dellacherie, C. and Meyer, P.-A. (1978). *Probabilities and Potential*, volume A. North-Holland.
[13] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer, 2nd edition.
[14] Diaconis, P. and Freedman, D. (1981). Partial exchangeability and sufficiency. In J. K. Ghosh and J. Roy, editors, *Proc. Indian Stat. Inst. Golden Jubilee Int'l Conf. Stat.: Applications and New Directions*, pages 205–236. Indian Statistical Institute.
[15] Diaconis, P. and Freedman, D. (2004). The Markov moment problem and de Finetti's theorem. Parts I and II. *Math. Z.*, **247**(1), 183–199 and 201–212.
[16] Diaconis, P. and Freedman, D. A. (1990). Cauchy's equation and de Finetti's theorem. *Scand. J. Stat.*, **17**(3), pp. 235–249.
[17] Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica, Serie VII*, **28**, 33–61.
[18] Favaro, S., Lijoi, A., and Prünster, I. (2012). Conditional formulae for Gibbs-type exchangeable random partitions. *Ann. Appl. Probab.* To appear.
[19] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**(2).
[20] Fligner, M. A. and Verducci, J. S. (1986). Distance based ranking models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **48**(3), 359–369.
[21] Florens, J. P., Mourchart, M., and Rolin, J. M. (1990). *Elements of Bayesian Statistics*. Marcel Dekker.
[22] Fremlin, D. H. (2000–2006). *Measure Theory*, volume I–V. Torres Fremlin.
[23] Ghosal, S. (2010). Dirichlet process, related priors and posterior asymptotics. In N. L. Hjort *et al.*, editors, *Bayesian Nonparametrics*. Cambridge University Press.
[24] Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.*, **35**(2), 697–723.
[25] Ghosh, J. K. and Ramamoorthi, R. V. (2002). *Bayesian Nonparametrics*. Springer.
[26] Hjort, N., Holmes, C., Müller, P., and Walker, S., editors (2010). *Bayesian Nonparametrics*. Cambridge University Press.
[27] James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Annals of Statistics*, **33**(4), 1771–1799.
[28] Kallenberg, O. (2001). *Foundations of Modern Probability*. Springer, 2nd edition.
[29] Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. Springer.

[30] Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Springer.
[31] Kechris, A. S. and Miller, B. D. (2004). *Topics in Orbit Equivalence*. Number 1852 in Lecture Notes in Mathematics. Springer.
[32] Kerov, S. V., Okounkov, A., and Olshanski, G. (1998). The boundary of Young graph with Jack edge multiplicities. *Internat. Math. Res. Notices*, **1998**, 173–199.
[33] Kerov, S. V., Olshanski, G. I., and Vershik, A. M. (2004). Harmonic analysis on the infinite symmetric group. *Invent. Math.*, **158**, 551–642.
[34] Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.*, **27**(2), 562–588.
[35] Küchler, U. and Lauritzen, S. L. (1989). Exponential families, extreme point models and minimal space-time invariant functions for stochastic processes with stationary and independent increments. *Scand. J. Stat.*, **16**, 237–261.
[36] Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. Lecture Notes in Statistics. Springer.
[37] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In N. L. Hjort *et al.*, editors, *Bayesian Nonparametrics*. Cambridge University Press.
[38] Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B*, **96**, 933–957.
[39] Lovász, L. and Szegedy, B. (2007). Szemerédi's lemma for the analyst. *Geometric And Functional Analysis*, **17**(1), 252–270.
[40] Meilă, M. and Bao, L. (2008). Estimation and clustering with infinite rankings. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[41] Orbanz, P. (2011). Projective limit random probabilities on Polish spaces. *Electron. J. Stat.*, **5**, 1354–1373.
[42] Parthasarathy, K. R. (1967). *Probability measures on metric spaces*. Academic Press.
[43] Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer.
[44] Pollard, D. (1977). Induced weak convergence and random measures. *Probab. Theory Related Fields*, **37**, 321–328.
[45] Schervish, M. J. (1995). *Theory of Statistics*. Springer.
[46] Schweinsberg, J. (2010). The number of small blocks in exchangeable random partitions. *ALEA Lat. Am. J. Probab. Math. Stat.*, **7**, 217–242.
[47] van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, **36**(3), 1435–1463.
[48] von Weizsäcker, H. and Winkler, G. (1979). Integral representation in the set of solutions of a generalized moment problem. *Math. Ann.*, **246**(1), 23–32.
[49] Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **61**(3), 485–527.
[50] Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.*, **28**, 532–552.